# BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.**

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 09:50:29 ; Search time 2238.85 Seconds
(without alignments)
174.827 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 297742

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 18.8 | 64.8 | 86 | 12 | CNS02IOD | AL199174 Tetraodon |
| c 2 | 16.4 | 56.6 | 96 | 12 | AZ786158 | AZ786158 2M0031E01 |
| 3 | 16.2 | 55.9 | 57 | 9 | AI561770 | AI561770 vv65b08.x |
| 4 | 16.2 | 55.9 | 100 | 9 | AI718731 | AI718731 as95f07.x |
| 5 | 16 | 55.2 | 77 | 12 | BH252676 | BH252676 SALK_0137 |
| 6 | 16 | 55.2 | 90 | 9 | AI636240 | AI636240 tz93c02.x |
| 7 | 16 | 55.2 | 96 | 12 | AZ923281 | AZ923281 4908.gf20 |
| c 8 | 15.4 | 53.1 | 66 | 12 | TA123H02P | AL463084 T. brucei |
| c 9 | 15.4 | 53.1 | 81 | 12 | AZ983583 | AZ983583 2M0264J16 |
| c 10 | 15.2 | 52.4 | 100 | 10 | H13996 | H13996 EST00022 Ch |
| 11 | 15 | 51.7 | 89 | 10 | BE545968 | BE545968 601069329 |
| c 12 | 14.8 | 51.0 | 25 | 12 | AZ993079 | AZ993079 2M0277P20 |
| 13 | 14.8 | 51.0 | 60 | 10 | BE871815 | BE871815 601447803 |
| 14 | 14.8 | 51.0 | 69 | 10 | BM128463 | BM128463 if15c05.x |
| c 15 | 14.8 | 51.0 | 72 | 12 | TA185E01P | AL474065 T. brucei |
| c 16 | 14.8 | 51.0 | 73 | 12 | AZ775013 | AZ775013 2M0004K20 |
| c 17 | 14.8 | 51.0 | 74 | 9 | AI819570 | AI819570 wj91g08.x |
| c 18 | 14.8 | 51.0 | 76 | 9 | AA975663 | AA975663 oq63f07.s |
| 19 | 14.8 | 51.0 | 76 | 10 | BI963697 | BI963697 ie63a09.x |
| 20 | 14.8 | 51.0 | 76 | 12 | AZ657549 | AZ657549 1M0533L18 |
| c 21 | 14.8 | 51.0 | 81 | 9 | AA922439 | AA922439 om63a07.s |
| 22 | 14.8 | 51.0 | 82 | 9 | AL588462 | AL588462 AL588462 |
| c 23 | 14.8 | 51.0 | 85 | 10 | BG370316 | BG370316 nai30d07. |
| c 24 | 14.8 | 51.0 | 86 | 10 | BG271746 | BG271746 nai53b10. |
| c 25 | 14.8 | 51.0 | 86 | 10 | BG271761 | BG271761 nai53d03. |
| c 26 | 14.8 | 51.0 | 87 | 10 | BG271786 | BG271786 nai53g06. |
| c 27 | 14.8 | 51.0 | 88 | 10 | BG272672 | BG272672 nah35d04. |
| c 28 | 14.8 | 51.0 | 88 | 10 | BG272928 | BG272928 nai56b09. |
| c 29 | 14.8 | 51.0 | 91 | 12 | AZ831658 | AZ831658 2M0111F03 |
| 30 | 14.8 | 51.0 | 92 | 10 | BF131941 | BF131941 601820909 |
| 31 | 14.8 | 51.0 | 93 | 12 | BH173468 | BH173468 0104C35 m |
| 32 | 14.8 | 51.0 | 94 | 10 | D11790 | D11790 HUMHM01G01 |
| 33 | 14.8 | 51.0 | 94 | 10 | D12351 | D12351 HUM00TB036 |
| 34 | 14.8 | 51.0 | 94 | 10 | D12363 | D12363 HUM00TB048 |
| 35 | 14.8 | 51.0 | 98 | 9 | AW059529 | AW059529 HuTH.bsst |
| c 36 | 14.8 | 51.0 | 100 | 9 | AI015946 | AI015946 ov30a01.x |
| 37 | 14.8 | 51.0 | 100 | 9 | AI299631 | AI299631 qm91f11.x |
| 38 | 14.6 | 50.3 | 52 | 10 | BG236504 | BG236504 nai44c06. |
| c 39 | 14.6 | 50.3 | 64 | 9 | AI139668 | AI139668 qc28h07.x |
| 40 | 14.6 | 50.3 | 64 | 12 | AZ808107 | AZ808107 2M0071O24 |
| 41 | 14.6 | 50.3 | 66 | 12 | AZ440181 | AZ440181 1M0231E10 |
| c 42 | 14.6 | 50.3 | 73 | 10 | BI865610 | BI865610 ft22h02.y |
| 43 | 14.6 | 50.3 | 79 | 10 | D11626 | D11626 HUM0C13G12 |
| c 44 | 14.6 | 50.3 | 94 | 9 | AA262376 | AA262376 zs26h02.r |
| c 45 | 14.6 | 50.3 | 95 | 10 | BI963950 | BI963950 ie63a09.y |

ALIGNMENTS

RESULT 1
CNS02IOD
LOCUS CNS02IOD 86 bp DNA linear GSS 13-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 141P20 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL199174
VERSION AL199174.1 GI:7837325
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 86)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 86)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 86)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .86
/organism="Tetraodon nigroviridis"

```
                     /db_xref="taxon:99883"
                     /clone="141P20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG141DH10LP1-end : T7"
BASE COUNT       12 a     12 c     26 g     34 t      2 others
ORIGIN


  Query Match             64.8%;  Score 18.8;  DB 12;  Length 86;
  Best Local Similarity   50.0%;  Pred. No. 4.4e+03;
  Matches   11;  Conservative    9;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 uucuuuuuguaagcccuagggg 27
           ::|::::::|:| | ||:|||||
Db      62 TTCTTTTTGTAGGGCCTAGGGG 83


RESULT   2
AZ786158/c
LOCUS       AZ786158                 96 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0031E01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0031E01 R, DNA sequence.
ACCESSION   AZ786158
VERSION     AZ786158.1  GI:12923638
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 96)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0031  row: E  column: 01
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 96.
FEATURES             Location/Qualifiers
     source          1. .96
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0031E01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       39 a     25 c     19 g     13 t
ORIGIN


  Query Match             56.6%;  Score 16.4;  DB 12;  Length 96;
  Best Local Similarity   42.3%;  Pred. No. 3.3e+04;
  Matches   11;  Conservative    9;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagcccuaggggcu 29
           | ::|::::::| | || |:  ||||:
Db      53 GTTTCTTTTTGGAGGCACTCAGGGCT 28


RESULT   3
AI561770
LOCUS       AI561770                 57 bp    mRNA    linear   EST 25-MAR-1999
DEFINITION  vv65b08.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            IMAGE:1227255 3', mRNA sequence.
ACCESSION   AI561770
VERSION     AI561770.1  GI:4513115
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 57)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:652847
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 51.
FEATURES             Location/Qualifiers
     source          1. .57
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1227255"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       17 a      9 c     12 g     19 t
ORIGIN


  Query Match             55.9%;  Score 16.2;  DB 9;  Length 57;
  Best Local Similarity   37.9%;  Pred. No. 4.3e+04;
```

```
  Matches   11;  Conservative   10;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggggcu 29
          : :|| :| :::: :|| || : |||| :
Db     25 TTTGAATCCTTTTCTAATCCATGGGGGGT 53


RESULT   4
AI718731
LOCUS       AI718731                 100 bp    mRNA    linear   EST 10-JUN-1999
DEFINITION  as95f07.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone
            IMAGE:2353381 3', mRNA sequence.
ACCESSION   AI718731
VERSION     AI718731.1  GI:5035987
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .100
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2353381"
                     /clone_lib="Barstead aorta HPLRB6"
                     /sex="male"
                     /dev_stage="adult, age 64"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: aorta; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [5' AATTCGGATCGAAC 3' and 5' GTTGGATCGG 3'], digested
                     with Not I and cloned into the Not I and Eco RI sites of
                     the modified pT7T3 vector. Library constructed by Bob
                     Barstead."
BASE COUNT       28 a     12 c     28 g     32 t
ORIGIN


  Query Match             55.9%;  Score 16.2;  DB 9;  Length 100;
  Best Local Similarity   44.8%;  Pred. No. 3.9e+04;
  Matches   13;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggggcu 29
          :|:||:  ::  :|:|||  | || |||:
Db     12 TATGATGTTTAATGTAAGATCCAGTGGCT 40


RESULT   5
BH252676
LOCUS       BH252676                 77 bp    DNA     linear   GSS 28-NOV-2001
DEFINITION  SALK_013733 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_013733, DNA sequence.
ACCESSION   BH252676
VERSION     BH252676.1  GI:17139654
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 77)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Heller,C., Kim,C.J., Jeske,A., Koesema,E., Meyers,M.C., Parker
            ,H., Prednis,L., Shinn,P., Stevenson,D.K., Zimmerman,J. and Ecker
            ,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .77
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_013733"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT       19 a     19 c     18 g     19 t      2 others
ORIGIN


  Query Match             55.2%;  Score 16;  DB 12;  Length 77;
  Best Local Similarity   43.8%;  Pred. No. 4.9e+04;
  Matches    7;  Conservative    9;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 augauucuuuuuguaa 17
            |:||::|:::::|:||
Db     10 ATGATTCTTTTTGTAA 25


RESULT   6
AI636240
LOCUS       AI636240                 90 bp    mRNA    linear   EST 16-DEC-1999
DEFINITION  tz93c02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2296130 3',
            mRNA sequence.
ACCESSION   AI636240
VERSION     AI636240.1  GI:4687570
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 90)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
```

```
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
           Insert Length: 1423   Std Error: 0.00
           Seq primer: -40UP from Gibco
           High quality sequence stop: 57.
FEATURES             Location/Qualifiers
     source          1. .90
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2296130"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT       21 a     17 c     25 g     27 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 9;  Length 90;
  Best Local Similarity   37.5%;  Pred. No. 4.7e+04;
  Matches    9;  Conservative   10;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 uucuuuuuguaagcccuaggggcu 29
           ::|:::::  :  | ||:||||| :
Db       4 TTCTTTTTTTTGGACCTAGGGGAT 27


RESULT   7
AZ923281
LOCUS       AZ923281                 96 bp    DNA     linear   GSS 01-APR-2001
DEFINITION  4908.gf20g16.s1 Saccharomyces cariocanus UFRJ 50791 Saccharomyces
            cariocanus genomic clone 4908.gf20g16.s1, DNA sequence.
ACCESSION   AZ923281
VERSION     AZ923281.1  GI:13494179
KEYWORDS    GSS.
SOURCE      Saccharomyces cariocanus.
  ORGANISM  Saccharomyces cariocanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 96)
  AUTHORS   Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
            ,W.R., Waterston,R.H. and Johnston,M.
  TITLE     Surveying Saccharomyces genomes to identify functional elements by
            comparative DNA sequence analysis
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Johnston M
            Department of Genetics
            Washington University Medical School
            Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
            Tel: 314 362 2735
            Fax: 314 362 7855
            Email: mj@genetics.wustl.edu
            Class: random plasmid subclone.
FEATURES             Location/Qualifiers
     source          1. .96
                     /organism="Saccharomyces cariocanus"
                     /strain="UFRJ 50791"
                     /db_xref="taxon:114526"
                     /clone="4908.gf20g16.s1"
                     /clone_lib="Saccharomyces cariocanus UFRJ 50791"
                     /note="Random genomic sequence "
BASE COUNT       27 a     20 c     16 g     33 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 12;  Length 96;
  Best Local Similarity   41.7%;  Pred. No. 4.7e+04;
  Matches   10;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 uucuuuuuguaagcccuaggggcu 29
           :: ::::: :|| ||  ||||||:
Db      73 TTTTTTTTCTAAACCAAAGGGGCT 96


RESULT   8
TA123H02P/c
LOCUS       TA123H02P                66 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 123h02, forward sequence,
            genomic survey sequence.
ACCESSION   AL463084
VERSION     AL463084.1  GI:11833690
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 66)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nhl@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .66
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="123h02"
BASE COUNT       29 a      8 c      9 g     20 t
ORIGIN


  Query Match             53.1%;  Score 15.4;  DB 12;  Length 66;
  Best Local Similarity   36.0%;  Pred. No. 8.3e+04;
  Matches    9;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 uaugauucuuuuuguaagcccuagg 25
           :|:||:: :::::   | |||:| |
Db      30 TATGATTTTTTTTCAGAACCCTAAG 6


RESULT   9
AZ983583/c
LOCUS       AZ983583                 81 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0264J16R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0264J16 R, DNA sequence.
ACCESSION   AZ983583
VERSION     AZ983583.1  GI:13854810
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 81)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0264  row: J  column: 16
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 81.
FEATURES              Location/Qualifiers
     source           1. .81
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0264J16"
                      /clone_lib="Mouse 10kb plasmid UUGC2M library"
                      /sex="Female"
                      /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (female) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       28 a     10 c     33 g     10 t
ORIGIN


  Query Match             53.1%;  Score 15.4;  DB 12;  Length 81;
  Best Local Similarity   48.0%;  Pred. No. 8.1e+04;
  Matches   12;  Conservative    7;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      5 auucuuuuuguaagcccuaggggcu 29
          |::|::: :|   |||  ||||||:
Db     45 ATTCTTTCTGCTGGCCTCAGGGGCT 21


RESULT  10
H13996/c
LOCUS       H13996                   100 bp    mRNA    linear   EST 03-JUL-1995
DEFINITION  EST00022 Chromosome 19p12-p13.1 exon Homo sapiens cDNA clone C3-8
            5', mRNA sequence.
ACCESSION   H13996
VERSION     H13996.1  GI:888005
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   Li,Q.Y.
  TITLE     Chromosome 19p12-p13.1 exons
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Li QY
            Human Molecular Genetics
            Queen's Medical Centre
            Nottingham, NG7 2UH, UK
            Tel: 1159249924
            Fax: 1159709906
            Email: pdzqyl@pdn1.gene.nottingham.ac.uk
            Seq primer: SD2 : 5' ATC TCA GTG GTA TTT GTC AGC 3'.
FEATURES              Location/Qualifiers
     source           1. .100
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="19p12-p13.1"
                      /clone="C3-8"
                      /clone_lib="Chromosome 19p12-p13.1 exon"
                      /lab_host="E. coli DH5a"
                      /note="Vector: pAMP10; Exons were isolated from human
                      chromosome 19p12-p13.1 specific  cosmids from Lawrence
                      Livermore National Laboratory using a modification of  the
                      method of exon amplification (Proc. Natl. Acad. Sci. USA
                      88: 4005-4009,  1991). Amplified exons were cloned into
                      pAMP10 by uracial cloning  (GIBCOL BRL)."
BASE COUNT       22 a     35 c     24 g     19 t
ORIGIN


  Query Match             52.4%;  Score 15.2;  DB 10;  Length 100;
  Best Local Similarity   46.4%;  Pred. No. 9.2e+04;
  Matches   13;  Conservative    7;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      2 augauucuuuuuguaagcccuaggggcu 29
            | ||:  :::: |: ||  | ||||||:
Db     96 AAGATGATTTTGGTTAGGTCCAGGGGCT 69


RESULT  11
BE545968
LOCUS       BE545968                 89 bp    mRNA    linear   EST 09-AUG-2000
DEFINITION  601069329F2 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3455694 5',
            mRNA sequence.
ACCESSION   BE545968
VERSION     BE545968.1  GI:9774613
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 89)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8442  row: i  column: 07
            High quality sequence stop: 89.
FEATURES              Location/Qualifiers
     source           1. .89
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
```

```
                     /clone="IMAGE:3455694"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT       41 a     14 c     12 g     22 t
ORIGIN


  Query Match             51.7%;  Score 15;  DB 10;  Length 89;
  Best Local Similarity   39.1%;  Pred. No. 1.1e+05;
  Matches    9;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccua 23
          :|:||  |:::::  :| || |:|
Db     21 TATGACACTTTTTCTAGGCTCTA 43


RESULT  12
AZ993079/c
LOCUS       AZ993079                 25 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0277P20R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0277P20 R, DNA sequence.
ACCESSION   AZ993079
VERSION     AZ993079.1  GI:13864306
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0277  row: P  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 25.
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0277P20"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       4 a      8 c      6 g      7 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 12;  Length 25;
  Best Local Similarity   72.2%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     11 uuuguaagcccuaggggc 28
          :::| |||||| ||||||
Db     21 TTTGCAAGCCCAAGGGGC 4


RESULT  13
BE871815
LOCUS       BE871815                 60 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601447803F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3851880 5',
            mRNA sequence.
ACCESSION   BE871815
VERSION     BE871815.1  GI:10320591
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9573  row: e  column: 01
            High quality sequence stop: 60.
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3851880"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT       16 a     11 c     10 g     23 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 10;  Length 60;
  Best Local Similarity   38.9%;  Pred. No. 1.4e+05;
  Matches    7;  Conservative    9;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 augauucuuuuuguaagc 19
          |:||:: ::::: :||||
Db     10 ATGATTATTTTTCTAAGC 27
```

```
RESULT  14
BM128463
LOCUS       BM128463                 69 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  if15c05.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 3', mRNA sequence.
ACCESSION   BM128463
VERSION     BM128463.1  GI:17123015
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu).
FEATURES             Location/Qualifiers
     source          1. .69
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
BASE COUNT        8 a     10 c     19 g     32 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 10;  Length 69;
  Best Local Similarity   30.8%;  Pred. No. 1.4e+05;
  Matches    8;  Conservative   11;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggg 26
          : :  :: ::::: :  ||||:||||
Db      8 TTTTTTTTTTTTTCTGGGCCCTAGGG 33


RESULT  15
TA185E01P/c
LOCUS       TA185E01P                72 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 185e01, forward sequence,
            genomic survey sequence.
ACCESSION   AL474065
VERSION     AL474065.1  GI:11840836
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 72)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .72
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="185e01"
BASE COUNT       14 a     16 c     25 g     17 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 12;  Length 72;
  Best Local Similarity   61.1%;  Pred. No. 1.4e+05;
  Matches   11;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      12 uuguaagcccuaggggcu 29
           :: :||||| :||||||:
Db      66 TTCTAAGCCATAGGGGCT 49


Search completed: September 13, 2002, 12:36:31
Job time: 9962 sec
```

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 09:54:15 ; Search time 63.83 Seconds
(without alignments)
111.599 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 613726

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| | Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| c | 1 | 15.2 | 52.4 | 25 | 4 | US-08-943-731-336 | Sequence 336, App |
| c | 2 | 15.2 | 52.4 | 33 | 1 | US-08-667-079B-5 | Sequence 5, Appli |
| | 3 | 15.2 | 52.4 | 90 | 1 | US-08-123-702-10 | Sequence 10, Appl |
| c | 4 | 14.8 | 51.0 | 35 | 6 | 5422260-12 | Patent No. 5422260 |
| c | 5 | 14.8 | 51.0 | 36 | 4 | US-09-440-001-1 | Sequence 1, Appli |
| | 6 | 14 | 48.3 | 37 | 1 | US-08-049-264C-55 | Sequence 55, Appl |
| | 7 | 14 | 48.3 | 37 | 1 | US-08-476-562-55 | Sequence 55, Appl |
| | 8 | 14 | 48.3 | 37 | 1 | US-08-479-723A-55 | Sequence 55, Appl |
| | 9 | 14 | 48.3 | 37 | 5 | PCT-US94-04310-55 | Sequence 55, Appl |
| c | 10 | 14 | 48.3 | 44 | 1 | US-08-049-264C-54 | Sequence 54, Appl |
| c | 11 | 14 | 48.3 | 44 | 1 | US-08-476-562-54 | Sequence 54, Appl |
| c | 12 | 14 | 48.3 | 44 | 1 | US-08-479-723A-54 | Sequence 54, Appl |
| c | 13 | 14 | 48.3 | 44 | 5 | PCT-US94-04310-54 | Sequence 54, Appl |
| c | 14 | 13.8 | 47.6 | 36 | 4 | US-09-440-001-3 | Sequence 3, Appli |
| | 15 | 13.6 | 46.9 | 44 | 2 | US-08-343-443B-39 | Sequence 39, Appl |
| | 16 | 13.4 | 46.2 | 26 | 4 | US-09-247-190-37 | Sequence 37, Appl |
| | 17 | 13.4 | 46.2 | 41 | 1 | US-08-468-220-28 | Sequence 28, Appl |
| | 18 | 13.4 | 46.2 | 41 | 2 | US-08-468-698-28 | Sequence 28, Appl |
| | 19 | 13.4 | 46.2 | 41 | 4 | US-08-194-664A-28 | Sequence 28, Appl |
| | 20 | 13.4 | 46.2 | 41 | 5 | PCT-US94-01553A-28 | Sequence 28, Appl |
| | 21 | 13.4 | 46.2 | 41 | 5 | PCT-US95-10426-28 | Sequence 28, Appl |
| c | 22 | 13.4 | 46.2 | 51 | 1 | US-08-328-152A-11 | Sequence 11, Appl |
| c | 23 | 13.4 | 46.2 | 60 | 3 | US-08-478-097A-32 | Sequence 32, Appl |
| c | 24 | 13.2 | 45.5 | 19 | 3 | US-08-532-896-53 | Sequence 53, Appl |
| | 25 | 13.2 | 45.5 | 27 | 3 | US-09-106-182-17 | Sequence 17, Appl |
| | 26 | 13.2 | 45.5 | 27 | 3 | US-09-106-182-23 | Sequence 23, Appl |
| | 27 | 13.2 | 45.5 | 27 | 4 | US-09-227-357-4 | Sequence 4, Appli |
| | 28 | 13.2 | 45.5 | 36 | 2 | US-08-882-083-7 | Sequence 7, Appli |
| | 29 | 13.2 | 45.5 | 36 | 2 | US-08-558-107-7 | Sequence 7, Appli |
| | 30 | 13.2 | 45.5 | 36 | 3 | US-09-243-539-7 | Sequence 7, Appli |
| c | 31 | 13.2 | 45.5 | 87 | 4 | US-09-364-539-128 | Sequence 128, App |
| c | 32 | 13 | 44.8 | 22 | 1 | US-08-647-584-118 | Sequence 118, App |
| c | 33 | 13 | 44.8 | 53 | 2 | US-08-486-969-46 | Sequence 46, Appl |
| c | 34 | 13 | 44.8 | 53 | 2 | US-08-687-865A-10 | Sequence 10, Appl |
| c | 35 | 13 | 44.8 | 53 | 4 | US-09-043-711-10 | Sequence 10, Appl |
| | 36 | 13 | 44.8 | 55 | 2 | US-08-687-865A-11 | Sequence 11, Appl |
| | 37 | 13 | 44.8 | 55 | 4 | US-09-043-711-11 | Sequence 11, Appl |
| c | 38 | 13 | 44.8 | 57 | 1 | US-08-192-300-14 | Sequence 14, Appl |
| | 39 | 13 | 44.8 | 67 | 4 | US-09-275-850-319 | Sequence 319, App |
| c | 40 | 13 | 44.8 | 70 | 4 | US-09-364-380-12 | Sequence 12, Appl |
| c | 41 | 13 | 44.8 | 72 | 1 | US-08-009-265-41 | Sequence 41, Appl |
| c | 42 | 13 | 44.8 | 78 | 1 | US-08-290-592E-40 | Sequence 40, Appl |
| c | 43 | 13 | 44.8 | 78 | 5 | PCT-US96-09448-40 | Sequence 40, Appl |
| | 44 | 13 | 44.8 | 97 | 1 | US-08-210-222-11 | Sequence 11, Appl |
| | 45 | 12.8 | 44.1 | 24 | 1 | US-08-508-778A-6 | Sequence 6, Appli |

ALIGNMENTS

```
RESULT   1
US-08-943-731-336/c
; Sequence 336, Application US/08943731
; Patent No. 6265157
;  GENERAL INFORMATION:
;    APPLICANT:  PROCKOP, DARWIN J.
;    APPLICANT:  SPOTILA, LORETTA D.
;    APPLICANT:  DELTAS, CONSTANTINOS D.
;    APPLICANT:  SEREDA, LARISA
;    APPLICANT:  LARSON, ANDREA W.
;    APPLICANT:  PACK, MICHAEL
;    APPLICANT:  COLIGE, ALAIN
;    APPLICANT:  EARLY, JAMES
;    APPLICANT:  KORKKO, JARMO
;    APPLICANT:  ALA-KOKKO, LEENA, et al.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;    NUMBER OF SEQUENCES:  666
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;      STREET:  FLR.
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/943,731
;      FILING DATE:  03-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/212,322
;      FILING DATE:  14-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/803,628
;      FILING DATE:  03-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOYLE LEARY Ph.D., KATHRYN
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9598-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  336:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  25 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-943-731-336


  Query Match             52.4%;  Score 15.2;  DB 4;  Length 25;
  Best Local Similarity   45.0%;  Pred. No. 1.4e+02;
  Matches    9;  Conservative    8;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 auucuuuuuguaagcccuag 24
            |::|: :::|: ||||| |
Db      22 ATTCTCTTTGTGAGCCCTTG 3


RESULT   2
US-08-667-079B-5/c
; Sequence 5, Application US/08667079B
; Patent No. 5789171
;  GENERAL INFORMATION:
;    APPLICANT:  Mark S. Smeltzer
;    TITLE OF INVENTION:  Use of cna, fnbA, fnbB, and hlb Gene Probes for the Strain-Sp
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Benjamin Aaron Adler, MCGREGOR & ADLER, P.C.
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/667,079B
;      FILING DATE:  June 20, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler, Benjamin Aaron
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5886
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    ORIGINAL SOURCE:
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
US-08-667-079B-5


  Query Match             52.4%;  Score 15.2;  DB 1;  Length 33;
  Best Local Similarity   39.3%;  Pred. No. 1.5e+02;
  Matches   11;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 augauucuuuuuguaagcccuaggggcu 29
            |:||:: :::: |:||   |   ||||:
Db      32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT   3
US-08-123-702-10
; Sequence 10, Application US/08123702
; Patent No. 5604131
;  GENERAL INFORMATION:
;    APPLICANT:  Wadsworth, Samuel
;    APPLICANT:  Snyder, Benjamin
;    APPLICANT:  Reddy, Vermuri, B.
;    APPLICANT:  Wei, Chamer
;    TITLE OF INVENTION:  A cDNA Genomic Hybrid Sequence Encoding APP770
; Patent No. 5604131
;    TITLE OF INVENTION:  Containing a Genomic DNA Insert of the KI and OX-2 Regions
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  GA
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/123,702
;      FILING DATE:  17-SEPT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  TSI121
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)-873-8794
;      TELEFAX:  (404)-873-8795
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  90 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-123-702-10


  Query Match             52.4%;  Score 15.2;  DB 1;  Length 90;
  Best Local Similarity   35.0%;  Pred. No. 1.8e+02;
  Matches    7;  Conservative   10;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 ugauucuuuuuguaagcccu 22
            :||::|:::::|:  || |:
Db      56 TGATTCTTTTTGTGTGCTCT 75


RESULT   4
5422260-12/c
;Patent No. 5422260
;    APPLICANT: KAUFMAN, RANDAL J.;PITTMAN, DEBRA D.;TOOLE,JOHN J.
;    TITLE OF INVENTION: HUMAN FACTOR VIII: C MUTEINS
;    NUMBER OF SEQUENCES: 15
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/883,936
;      FILING DATE: 15-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 279,485
;      FILING DATE: 02-DEC-1988; 09-DEC-1986
```

```
;       APPLICATION NUMBER: 939,658
;       FILING DATE: 09-DEC-1986
;       APPLICATION NUMBER: 932,767
;       FILING DATE: 18-NOV-1986
;       APPLICATION NUMBER: 868,410
;       FILING DATE: 29-MAY-1986
;SEQ ID NO:12:
;       LENGTH: 35
5422260-12


  Query Match             51.0%;  Score 14.8;  DB 6;  Length 35;
  Best Local Similarity   42.3%;  Pred. No. 2.3e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagcccuaggggcu 29
           | ::  ::::| ||||  : |||||:
Db      35 GTTTTCTTTTGAAAGCTTTTGGGGCT 10


RESULT   5
US-09-440-001-1/c
; Sequence 1, Application US/09440001
; Patent No. 6174696
; GENERAL INFORMATION:
;  APPLICANT:  Seman, Leo J.
;  TITLE OF INVENTION:  A METHOD FOR THE DETERMINATION OF HOMOCYSTEINE
;  FILE REFERENCE:  09/440,001
;  CURRENT APPLICATION NUMBER: US/09/440,001
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  60/108,099
;  PRIOR FILING DATE:  1998-11-12
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 36
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: Oligonucleotide primer
US-09-440-001-1


  Query Match             51.0%;  Score 14.8;  DB 4;  Length 36;
  Best Local Similarity   38.5%;  Pred. No. 2.3e+02;
  Matches   10;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 uaugauucuuuuuguaagcccuaggg 26
           :|: |  |:::::|:  ||| :| ||
Db      33 TATCAAGCTTTTTGTCCGCCATATGG 8


RESULT   6
US-08-049-264C-55
; Sequence 55, Application US/08049264C
; Patent No. 5518901
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,264C
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-049-264C-55


  Query Match             48.3%;  Score 14;  DB 1;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 5.4e+02;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 uucuuuuuguaagcccuagggg 27
           :: ::::: :|| |||  ||||
Db       8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT   7
US-08-476-562-55
; Sequence 55, Application US/08476562
; Patent No. 5688669
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,562
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/049,264
;      FILING DATE:  April 19, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-476-562-55


  Query Match             48.3%;  Score 14;  DB 1;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 5.4e+02;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuagggg 27
          :: ::::: :|| |||  ||||
Db      8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT   8
US-08-479-723A-55
; Sequence 55, Application US/08479723A
; Patent No. 5744306
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,723A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  05010.0061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligonucleotide
US-08-479-723A-55


  Query Match             48.3%;  Score 14;  DB 1;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 5.4e+02;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuagggg 27
          :: ::::: :|| |||  ||||
Db      8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT   9
PCT-US94-04310-55
; Sequence 55, Application PC/TUS9404310
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  74
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04310
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/049,264
;      FILING DATE:  19-APR-1993
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US94-04310-55


  Query Match             48.3%;  Score 14;  DB 5;  Length 37;
  Best Local Similarity   40.9%;  Pred. No. 5.4e+02;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuagggg 27
          :: ::::: :|| |||  ||||
Db      8 TTTTTTTTTTAAACCCGGGGGG 29


RESULT  10
US-08-049-264C-54/c
; Sequence 54, Application US/08049264C
; Patent No. 5518901
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/049,264C
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-049-264C-54
```

```
  Query Match             48.3%;  Score 14;  DB 1;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 5.6e+02;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuagggg 27
          :: ::::: :|| |||  ||||
Db     42 TTTTTTTTTTAAACCCGGGGGG 21


RESULT  11
US-08-476-562-54/c
; Sequence 54, Application US/08476562
; Patent No. 5688669
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,562
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/049,264
;      FILING DATE:  April 19, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  1313.004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-476-562-54


  Query Match             48.3%;  Score 14;  DB 1;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 5.6e+02;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuagggg 27
          :: ::::: :|| |||  ||||
Db     42 TTTTTTTTTTAAACCCGGGGGG 21


RESULT  12
US-08-479-723A-54/c
; Sequence 54, Application US/08479723A
; Patent No. 5744306
;  GENERAL INFORMATION:
;    APPLICANT:  Murtagh, James J.
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg., 127
;      STREET:  Peachtree Street N.E.
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,723A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perryman, David G.
;      REGISTRATION NUMBER:  33,438
;      REFERENCE/DOCKET NUMBER:  05010.0061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 688-0770
;      TELEFAX:  (404) 688-9880
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligonucleotide
US-08-479-723A-54


  Query Match             48.3%;  Score 14;  DB 1;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 5.6e+02;
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuagggg 27
          :: ::::: :|| |||  ||||
Db     42 TTTTTTTTTTAAACCCGGGGGG 21


RESULT  13
PCT-US94-04310-54/c
; Sequence 54, Application PC/TUS9404310
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS FOR NUCLEIC ACID DETECTION,
;    TITLE OF INVENTION:  SEQUENCING AND CLONING USING EXONUCLEASE
;    NUMBER OF SEQUENCES:  74
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04310
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/049,264
;      FILING DATE:  19-APR-1993
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US94-04310-54


  Query Match             48.3%;  Score 14;  DB 5;  Length 44;
  Best Local Similarity   40.9%;  Pred. No. 5.6e+02;
```

```
  Matches    9;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuagggg 27
          :: ::::: :|| |||  ||||
Db     42 TTTTTTTTTTAAACCCGGGGGG 21


RESULT  14
US-09-440-001-3/c
; Sequence 3, Application US/09440001
; Patent No. 6174696
; GENERAL INFORMATION:
;  APPLICANT:  Seman, Leo J.
;  TITLE OF INVENTION:  A METHOD FOR THE DETERMINATION OF HOMOCYSTEINE
;  FILE REFERENCE:  09/440,001
;  CURRENT APPLICATION NUMBER: US/09/440,001
;  CURRENT FILING DATE:  1999-11-12
;  PRIOR APPLICATION NUMBER:  60/108,099
;  PRIOR FILING DATE:  1998-11-12
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 36
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: Oligonucleotide primer
US-09-440-001-3


  Query Match             47.6%;  Score 13.8;  DB 4;  Length 36;
  Best Local Similarity   40.0%;  Pred. No. 6.7e+02;
  Matches   10;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuagg 25
          :|: |  |:::::|:  ||||  ||
Db     28 TATCAAGCTTTTTGTCCGCCCCGGG 4


RESULT  15
US-08-343-443B-39
; Sequence 39, Application US/08343443B
; Patent No. 5968734
;  GENERAL INFORMATION:
;    APPLICANT:  Aurias, Alain
;    APPLICANT:  Delattre, Olivier
;    APPLICANT:  Desmaze, Chantal
;    APPLICANT:  Melot, Thomas
;    APPLICANT:  Peter, Martine
;    APPLICANT:  Ploougastel, Beatrice
;    APPLICANT:  Thomas, Gilles
;    APPLICANT:  Zucman, Jessica
;    TITLE OF INVENTION:  NUCLEIC ACID CORRESPONDING TO A GENE OF
;    TITLE OF INVENTION:  CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
;    TITLE OF INVENTION:  TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
;    TITLE OF INVENTION:  TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
;    TITLE OF INVENTION:  TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  129
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Weiser & Associates
;      STREET:  230 South Fifteenth Street
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AEDIT 1.0 DOS text editor
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/343,443B
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR93/00494
;      FILING DATE:  19-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/06123
;      FILING DATE:  20-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weiser, Gerard J.
;      REGISTRATION NUMBER:  19,763
;      REFERENCE/DOCKET NUMBER:  989.6121P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-875-8383
;      TELEFAX:  215-875-8394
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-343-443B-39


  Query Match             46.9%;  Score 13.6;  DB 2;  Length 44;
  Best Local Similarity   35.7%;  Pred. No. 8.6e+02;
  Matches   10;  Conservative    9;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggggc 28
          : :  ::|:: ::|:| | ||  | |||
Db     14 TGTTGTTCTTGTTGTAGGTCCAGGAGGC 41


Search completed: September 13, 2002, 12:37:52
Job time: 9817 sec
```

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 11:58:05 ; Search time 280.51 Seconds
(without alignments)
177.500 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 2046006

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 29 | 100.0 | 29 | 21 | AAA70828 | Molecular interact |
| 2 | 29 | 100.0 | 42 | 21 | AAA71123 | Molecular interact |
| 3 | 29 | 100.0 | 42 | 21 | AAA71131 | Molecular interact |
| 4 | 28 | 96.6 | 45 | 21 | AAA70824 | Molecular interact |
| 5 | 28 | 96.6 | 46 | 21 | AAA71087 | Molecular interact |
| 6 | 28 | 96.6 | 46 | 21 | AAA71096 | Molecular interact |
| 7 | 28 | 96.6 | 46 | 21 | AAA71099 | Molecular interact |
| 8 | 28 | 96.6 | 46 | 21 | AAA71100 | Molecular interact |
| 9 | 28 | 96.6 | 46 | 21 | AAA71104 | Molecular interact |
| 10 | 25.8 | 89.0 | 42 | 21 | AAA71113 | Molecular interact |
| 11 | 25.8 | 89.0 | 42 | 21 | AAA71118 | Molecular interact |
| 12 | 25.8 | 89.0 | 42 | 21 | AAA71126 | Molecular interact |
| 13 | 24.8 | 85.5 | 46 | 21 | AAA71085 | Molecular interact |
| 14 | 24.8 | 85.5 | 46 | 21 | AAA71103 | Molecular interact |
| 15 | 23.8 | 82.1 | 42 | 21 | AAA71114 | Molecular interact |
| 16 | 23.8 | 82.1 | 42 | 21 | AAA71119 | Molecular interact |
| 17 | 23.8 | 82.1 | 42 | 21 | AAA71127 | Molecular interact |
| 18 | 23.8 | 82.1 | 46 | 21 | AAA71094 | Molecular interact |
| 19 | 23.8 | 82.1 | 46 | 21 | AAA71110 | Molecular interact |
| 20 | 23.2 | 80.0 | 29 | 21 | AAA70829 | Molecular interact |
| 21 | 23.2 | 80.0 | 29 | 21 | AAA70830 | Molecular interact |
| 22 | 23.2 | 80.0 | 42 | 21 | AAA71115 | Molecular interact |
| 23 | 23.2 | 80.0 | 42 | 21 | AAA71116 | Molecular interact |
| 24 | 23.2 | 80.0 | 42 | 21 | AAA71120 | Molecular interact |
| 25 | 23.2 | 80.0 | 42 | 21 | AAA71121 | Molecular interact |
| 26 | 23.2 | 80.0 | 42 | 21 | AAA71128 | Molecular interact |
| 27 | 23.2 | 80.0 | 42 | 21 | AAA71129 | Molecular interact |
| 28 | 22.6 | 77.9 | 42 | 21 | AAA71124 | Molecular interact |
| 29 | 22.6 | 77.9 | 42 | 21 | AAA71132 | Molecular interact |
| 30 | 22.2 | 76.6 | 45 | 21 | AAA70825 | Molecular interact |
| 31 | 22.2 | 76.6 | 45 | 21 | AAA70826 | Molecular interact |
| 32 | 22.2 | 76.6 | 46 | 21 | AAA71088 | Molecular interact |
| 33 | 22.2 | 76.6 | 46 | 21 | AAA71089 | Molecular interact |
| 34 | 22.2 | 76.6 | 46 | 21 | AAA71090 | Molecular interact |
| 35 | 22.2 | 76.6 | 46 | 21 | AAA71105 | Molecular interact |
| 36 | 22.2 | 76.6 | 46 | 21 | AAA71106 | Molecular interact |
| 37 | 22.2 | 76.6 | 46 | 21 | AAA71107 | Molecular interact |
| 38 | 21.6 | 74.5 | 46 | 21 | AAA71093 | Molecular interact |
| 39 | 21.6 | 74.5 | 46 | 21 | AAA71095 | Molecular interact |
| 40 | 21.6 | 74.5 | 46 | 21 | AAA71109 | Molecular interact |
| 41 | 21.6 | 74.5 | 46 | 21 | AAA71111 | Molecular interact |
| 42 | 19.4 | 66.9 | 46 | 21 | AAA71084 | Molecular interact |
| 43 | 19.4 | 66.9 | 46 | 21 | AAA71098 | Molecular interact |
| 44 | 19.4 | 66.9 | 46 | 21 | AAA71102 | Molecular interact |
| 45 | 18.4 | 63.4 | 42 | 21 | AAA71117 | Molecular interact |

ALIGNMENTS

RESULT 1
AAA70828
ID AAA70828 standard; RNA; 29 BP.
XX
AC AAA70828;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #28.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Homo sapiens.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 5 A; 5 C; 7 G; 12 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.00092;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggcu 29
||||||||||||||||||||||||||||||
Db 1 uaugauucuuuuuguaagcccuaggggcu 29

RESULT 2
AAA71123
ID AAA71123 standard; DNA; 42 BP.
XX
AC AAA71123;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #129.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 9 A; 6 C; 9 G; 18 T; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 58.6%; Pred. No. 0.00096;
Matches 17; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggcu 29
:|:||::|:::::|:||||||:||||||:
Db 4 tatgattctttttgtaagccctaggggct 32

RESULT 3
AAA71131
ID AAA71131 standard; RNA; 42 BP.
XX
AC AAA71131;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #200.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 9 A; 6 C; 9 G; 18 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggcu 29
||||||||||||||||||||||||||||||
Db 4 uaugauucuuuuuguaagcccuaggggcu 32

RESULT 4
AAA70824
ID AAA70824 standard; RNA; 45 BP.
XX
AC AAA70824;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #24.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Homo sapiens.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Claim 220; Page 232; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 45 BP; 11 A; 6 C; 9 G; 19 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggc 28
|||||||||||||||||||||||||||||
Db 18 uaugauucuuuuuguaagcccuaggggc 45

RESULT 5
AAA71087
ID AAA71087 standard; DNA; 46 BP.
XX
AC AAA71087;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #110.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 60.7%; Pred. No. 0.0028;
Matches 17; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggc 28
:|:||::|:::::|:||||||:||||||
Db 19 tatgattctttttgtaagccctaggggc 46

RESULT 6
AAA71096
ID AAA71096 standard; DNA; 46 BP.
XX
AC AAA71096;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #119.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 60.7%; Pred. No. 0.0028;
Matches 17; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggc 28
:|:||::|:::::|:||||||:||||||
Db 19 tatgattctttttgtaagccctaggggc 46

RESULT 7
AAA71099
ID AAA71099 standard; DNA; 46 BP.
XX
AC AAA71099;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #122.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

```
PT   agricultural and industrial compounds -
XX
PS   Example 7; Figure 121; 405pp; English.
XX
CC   This invention describes a novel method for identifying compounds which
CC   modulate the activity of a target biomolecule. The method uses
CC   3-dimensional representations of the biomolecule and a library of
CC   compounds and comprises (a) identifying at least one molecular
CC   interaction site of the target RNA; (b) generating in silico a virtual
CC   library of compounds predicted or calculated to interact with the
CC   molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC   representations of the target RNA with members of the virtual library of
CC   compounds to generate a hierarchy of the compounds ranked in accordance
CC   with their respective ability to form physical interactions with the
CC   molecular interaction site. The method also describes (1) RNA comprising
CC   a joined sequence of at least 24 nucleotides but not more than 70
CC   nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC   forming a first side of a first double stranded (ds) region; (b) 2
CC   nucleotides forming a first side of an internal loop region; (c) 4
CC   nucleotides forming a first side of a second ds region; (d) 4 or 5
CC   nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC   second side of the second ds region; (f) 4 nucleotides forming a second
CC   side of the internal loop region; and (g) 3 nucleotides forming a second
CC   side of the first ds region; (2) a purified and isolated RNA fragment
CC   comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC   methods and products can be used for identifying agents which modulate
CC   the activity of biomolecules, particularly RNA. Such agents can be used
CC   as pharmaceutical, agricultural or industrial compounds.
XX
SQ   Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;


  Query Match             96.6%;  Score 28;  DB 21;  Length 46;
  Best Local Similarity   60.7%;  Pred. No. 0.0028;
  Matches   17;  Conservative   11;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 uaugauucuuuuuguaagcccuaggggc 28
           :|:||::|::::|:|||||||:||||||
Db      19 tatgattctttttgtaagccctaggggc 46


RESULT   8
AAA71100
ID   AAA71100 standard; DNA; 46 BP.
XX
AC   AAA71100;
XX
DT   27-APR-2001  (first entry)
XX
DE   Molecular interaction site DNA #123.
XX
KW   Modulator; identification; molecular interaction; virtual library; ss.
XX
OS   Unidentified.
XX
PN   WO9958947-A2.
XX
PD   18-NOV-1999.
XX
PF   12-MAY-1999;   99WO-US10361.
XX
PR   12-MAY-1998;   98US-0076404.
PR   12-MAY-1998;   98US-0085092.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Ecker DJ,  Griffey R,  Crooke ST,  Sampath R,  Swayze E,  Mohan V;
PI   Hofstadler S,  McNeil J;
XX
DR   WPI; 2000-086439/07.
XX
PT   Identifying compounds which modulate activity of target biomolecules,
PT   used to provide compounds which can be used as pharmacological,
```

```
PT   agricultural and industrial compounds -
XX
PS   Example 7; Figure 121; 405pp; English.
XX
CC   This invention describes a novel method for identifying compounds which
CC   modulate the activity of a target biomolecule. The method uses
CC   3-dimensional representations of the biomolecule and a library of
CC   compounds and comprises (a) identifying at least one molecular
CC   interaction site of the target RNA; (b) generating in silico a virtual
CC   library of compounds predicted or calculated to interact with the
CC   molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC   representations of the target RNA with members of the virtual library of
CC   compounds to generate a hierarchy of the compounds ranked in accordance
CC   with their respective ability to form physical interactions with the
CC   molecular interaction site. The method also describes (1) RNA comprising
CC   a joined sequence of at least 24 nucleotides but not more than 70
CC   nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC   forming a first side of a first double stranded (ds) region; (b) 2
CC   nucleotides forming a first side of an internal loop region; (c) 4
CC   nucleotides forming a first side of a second ds region; (d) 4 or 5
CC   nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC   second side of the second ds region; (f) 4 nucleotides forming a second
CC   side of the internal loop region; and (g) 3 nucleotides forming a second
CC   side of the first ds region; (2) a purified and isolated RNA fragment
CC   comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC   methods and products can be used for identifying agents which modulate
CC   the activity of biomolecules, particularly RNA. Such agents can be used
CC   as pharmaceutical, agricultural or industrial compounds.
XX
SQ   Sequence 46 BP; 11 A; 7 C; 9 G; 19 T; 0 other;


  Query Match             96.6%;  Score 28;  DB 21;  Length 46;
  Best Local Similarity   60.7%;  Pred. No. 0.0028;
  Matches   17;  Conservative   11;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 uaugauucuuuuuguaagcccuaggggc 28
           :|:||::|::::|:|||||||:||||||
Db      19 tatgattctttttgtaagccctaggggc 46


RESULT   9
AAA71104
ID   AAA71104 standard; RNA; 46 BP.
XX
AC   AAA71104;
XX
DT   27-APR-2001  (first entry)
XX
DE   Molecular interaction site RNA #180.
XX
KW   Modulator; identification; molecular interaction; virtual library; ss.
XX
OS   Unidentified.
XX
PN   WO9958947-A2.
XX
PD   18-NOV-1999.
XX
PF   12-MAY-1999;   99WO-US10361.
XX
PR   12-MAY-1998;   98US-0076404.
PR   12-MAY-1998;   98US-0085092.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Ecker DJ,  Griffey R,  Crooke ST,  Sampath R,  Swayze E,  Mohan V;
PI   Hofstadler S,  McNeil J;
XX
DR   WPI; 2000-086439/07.
XX
PT   Identifying compounds which modulate activity of target biomolecules,
PT   used to provide compounds which can be used as pharmacological,
```

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 11 A; 7 C; 9 G; 19 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggc 28
||||||||||||||||||||||||||||
Db 19 uaugauucuuuuuguaagcccuaggggc 46

RESULT 10
AAA71113
ID AAA71113 standard; RNA; 42 BP.
XX
AC AAA71113;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #189.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 U; 0 other;

Query Match 89.0%; Score 25.8; DB 21; Length 42;
Best Local Similarity 93.1%; Pred. No. 0.027;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggcu 29
|| |||||||||||||||||||| |||||
Db 4 uaagauucuuuuuguaagcccuacgggcu 32

RESULT 11
AAA71118
ID AAA71118 standard; DNA; 42 BP.
XX
AC AAA71118;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #124.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 T; 0 other;

Query Match 89.0%; Score 25.8; DB 21; Length 42;
Best Local Similarity 55.2%; Pred. No. 0.027;
Matches 16; Conservative 11; Mismatches 2; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggcu 29
:| ||::|:::::|:|||||||:| ||||:
Db 4 taagattctttttgtaagccctacgggct 32

RESULT 12
AAA71126
ID AAA71126 standard; RNA; 42 BP.
XX
AC AAA71126;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #195.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 12 A; 7 C; 6 G; 17 U; 0 other;

Query Match 89.0%; Score 25.8; DB 21; Length 42;
Best Local Similarity 93.1%; Pred. No. 0.027;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggcu 29
|| ||||||||||||||||||||| |||||
Db 4 uaagauucuuuuuguaagcccuacgggcu 32

RESULT 13
AAA71085
ID AAA71085 standard; DNA; 46 BP.
XX
AC AAA71085;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #108.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 12 A; 7 C; 9 G; 18 T; 0 other;

Query Match 85.5%; Score 24.8; DB 21; Length 46;
Best Local Similarity 57.1%; Pred. No. 0.076;
Matches 16; Conservative 10; Mismatches 2; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggc 28
:| ||::|:::::|:||||||:| ||||
Db 19 taagattctttttgtaagccctacgggc 46

RESULT 14
AAA71103
ID AAA71103 standard; RNA; 46 BP.
XX
AC AAA71103;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #179.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 12 A; 7 C; 9 G; 18 U; 0 other;

Query Match 85.5%; Score 24.8; DB 21; Length 46;
Best Local Similarity 92.9%; Pred. No. 0.076;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuaggggc 28
|| ||||||||||||||||||||| ||||
Db 19 uaagauucuuuuuguaagcccuacgggc 46

RESULT 15
AAA71114
ID AAA71114 standard; RNA; 42 BP.
XX
AC AAA71114;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #190.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 11 A; 8 C; 7 G; 16 U; 0 other;

Query Match 82.1%; Score 23.8; DB 21; Length 42;
Best Local Similarity 92.6%; Pred. No. 0.21;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 uaugauucuuuuuguaagcccuagggg 27
|| ||||||||||||||||||||||| |
Db 4 uaagauucuuuuuguaagcccuaggcg 30

Search completed: September 13, 2002, 13:23:15
Job time: 5110 sec

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 11:56:05 ; Search time 2058.64 Seconds
(without alignments)
294.791 Million cell updates/sec

Title: US-09-310-844C-24
Perfect score: 29
Sequence: 1 uaugauucuuuuuguaagcccuaggggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 843946

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| c 1 | 16 | 55.2 | 82 | 3 | AF144669 | AF144669 Patella v |
| c 2 | 15.2 | 52.4 | 33 | 6 | AR020509 | AR020509 Sequence |
| 3 | 15.2 | 52.4 | 90 | 6 | I36126 | I36126 Sequence 10 |
| c 4 | 15 | 51.7 | 40 | 6 | E49126 | E49126 Novel G pro |
| c 5 | 15 | 51.7 | 40 | 6 | E50836 | E50836 Novel G pro |
| c 6 | 15 | 51.7 | 87 | 3 | DDIDDKD | M59747 D.discoideu |
| c 7 | 14.8 | 51.0 | 35 | 6 | I08597 | I08597 Sequence 12 |
| c 8 | 14.8 | 51.0 | 36 | 6 | AR142180 | AR142180 Sequence |
| 9 | 14.6 | 50.3 | 51 | 6 | AX115517 | AX115517 Sequence |
| 10 | 14.6 | 50.3 | 51 | 10 | AF328713 | AF328713 Mus muscu |
| 11 | 14.6 | 50.3 | 63 | 9 | S63972 | S63972 IgH [CDR3 r |
| c 12 | 14.6 | 50.3 | 72 | 1 | ECOPKYD2 | M30729 E.coli/IS1 |
| 13 | 14.4 | 49.7 | 25 | 6 | AX042583 | AX042583 Sequence |
| 14 | 14.4 | 49.7 | 25 | 6 | AX043280 | AX043280 Sequence |
| c 15 | 14.2 | 49.0 | 35 | 6 | AX298174 | AX298174 Sequence |
| c 16 | 14.2 | 49.0 | 45 | 6 | I04390 | I04390 Sequence 25 |
| c 17 | 14.2 | 49.0 | 56 | 6 | AX247478 | AX247478 Sequence |
| 18 | 14.2 | 49.0 | 100 | 4 | AY045354 | AY045354 Loxodonta |
| 19 | 14.2 | 49.0 | 100 | 4 | AY045355 | AY045355 Loxodonta |
| 20 | 14.2 | 49.0 | 100 | 4 | AY045356 | AY045356 Loxodonta |
| 21 | 14.2 | 49.0 | 100 | 4 | AY045357 | AY045357 Loxodonta |
| 22 | 14.2 | 49.0 | 100 | 4 | AY045358 | AY045358 Loxodonta |
| 23 | 14.2 | 49.0 | 100 | 4 | AY045359 | AY045359 Loxodonta |
| 24 | 14.2 | 49.0 | 100 | 4 | AY045360 | AY045360 Loxodonta |
| 25 | 14.2 | 49.0 | 100 | 4 | AY045361 | AY045361 Loxodonta |
| 26 | 14.2 | 49.0 | 100 | 4 | AY045362 | AY045362 Loxodonta |
| 27 | 14.2 | 49.0 | 100 | 4 | AY045363 | AY045363 Loxodonta |
| c 28 | 14.2 | 49.0 | 100 | 5 | AF174501 | AF174501 Bufo hima |
| c 29 | 14.2 | 49.0 | 100 | 5 | AF174506 | AF174506 Bufo garg |
| 30 | 14 | 48.3 | 30 | 6 | AX113961 | AX113961 Sequence |
| c 31 | 14 | 48.3 | 30 | 6 | AX113963 | AX113963 Sequence |
| 32 | 14 | 48.3 | 37 | 6 | AR003420 | AR003420 Sequence |
| 33 | 14 | 48.3 | 37 | 6 | I21209 | I21209 Sequence 55 |
| 34 | 14 | 48.3 | 37 | 6 | I74476 | I74476 Sequence 55 |
| c 35 | 14 | 48.3 | 44 | 6 | AR003419 | AR003419 Sequence |
| c 36 | 14 | 48.3 | 44 | 6 | I21208 | I21208 Sequence 54 |
| c 37 | 14 | 48.3 | 44 | 6 | I74475 | I74475 Sequence 54 |
| c 38 | 14 | 48.3 | 78 | 14 | AF362846 | AF362846 HIV-1 iso |
| c 39 | 14 | 48.3 | 78 | 14 | AF362847 | AF362847 HIV-1 iso |
| c 40 | 14 | 48.3 | 78 | 14 | AF362849 | AF362849 HIV-1 iso |
| 41 | 14 | 48.3 | 100 | 9 | HUM0401I | M83920 Human MHC c |
| 42 | 14 | 48.3 | 100 | 9 | HUM0501I | M83921 Human MHC c |
| c 43 | 13.8 | 47.6 | 36 | 6 | AR142182 | AR142182 Sequence |
| 44 | 13.8 | 47.6 | 42 | 6 | AX017119 | AX017119 Sequence |
| c 45 | 13.8 | 47.6 | 42 | 6 | AX017120 | AX017120 Sequence |

ALIGNMENTS

RESULT 1
AF144669/c
LOCUS AF144669 82 bp DNA linear INV 04-AUG-1999
DEFINITION Patella vulgata antennapedia-like homeodomain protein HB3 gene, partial cds.
ACCESSION AF144669
VERSION AF144669.1 GI:5690267
KEYWORDS .
SOURCE common limpet.
ORGANISM Patella vulgata
Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda; Patelloidea; Patellidae; Patella.
REFERENCE 1 (bases 1 to 82)
AUTHORS de Rosa,R., Grenier,J.K., Andreeva,T., Cook,C.E., Adoutte,A., Akam,M., Carroll,S.B. and Balavoine,G.
TITLE Hox genes in brachiopods and priapulids and protostome evolution
JOURNAL Nature 399 (6738), 772-776 (1999)
MEDLINE 99318125
PUBMED 10391241
REFERENCE 2 (bases 1 to 82)
AUTHORS de Rosa,R., Lartillot,N. and Adoutte,A.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1999) Centre de Genetique Moleculaire, Avenue de

```
           la Terrasse, Gif-sur-Yvette 91198, France
FEATURES             Location/Qualifiers
     source          1. .82
                     /organism="Patella vulgata"
                     /db_xref="taxon:6465"
     mRNA            <1. .>82
                     /product="antennapedia-like homeodomain protein HB3"
     CDS             <1. .>82
                     /codon_start=2
                     /product="antennapedia-like homeodomain protein HB3"
                     /protein_id="AAD47008.1"
                     /db_xref="GI:5690268"
                     /translation="HFNRYLTRRRRIEVAHMLGLTERQIKI"
BASE COUNT       34 a     16 c     17 g     15 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 3;  Length 82;
  Best Local Similarity   45.8%;  Pred. No. 7.3e+03;
  Matches   11;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 uaugauucuuuuuguaagcccuag 24
           :|:   :|::: :|:|||||| ||
Db      76 TATTTGTCTTTCTGTAAGCCCGAG 53


RESULT   2
AR020509/c
LOCUS       AR020509                 33 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 5 from patent US 5789171.
ACCESSION   AR020509
VERSION     AR020509.1  GI:3975124
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Smeltzer,M.S.
  TITLE     Use of cna, fnba, fnbb, and hlb, gene probes for the
            strain-specific identification of Staphylococcus aureus
  JOURNAL   Patent: US 5789171-A 5 04-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="unknown"
BASE COUNT       12 a      8 c      7 g      6 t
ORIGIN


  Query Match             52.4%;  Score 15.2;  DB 6;  Length 33;
  Best Local Similarity   39.3%;  Pred. No. 1.8e+04;
  Matches   11;  Conservative    9;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 augauucuuuuuguaagcccuaggggcu 29
           |:||:: :::: |:||   |   ||||:
Db      32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT   3
I36126
LOCUS       I36126                   90 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 10 from patent US 5604131.
ACCESSION   I36126
VERSION     I36126.1  GI:2087350
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 90)
  AUTHORS   Wadsworth,S., Snyder,B., Reddy,V.B. and Wei,C.
  TITLE     cDNA-genomic DNA hybrid sequence encoding APP770 containing a
            genomic DNA insert of the KI and OX-2 regions
  JOURNAL   Patent: US 5604131-A 10 18-FEB-1997;
FEATURES             Location/Qualifiers
     source          1. .90
                     /organism="unknown"
BASE COUNT       18 a     23 c     20 g     29 t
ORIGIN


  Query Match             52.4%;  Score 15.2;  DB 6;  Length 90;
  Best Local Similarity   35.0%;  Pred. No. 1.7e+04;
  Matches    7;  Conservative   10;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 ugauucuuuuuguaagcccu 22
           :||:::|:::::|:  || |:
Db      56 TGATTCTTTTTGTGTGCTCT 75


RESULT   4
E49126/c
LOCUS       E49126                   40 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Novel G protein-conjugated receptor protein.
ACCESSION   E49126
VERSION     E49126.1  GI:18629263
KEYWORDS    JP 2001029083-A/4.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Takasaki,A., Matsumoto,M., Sugimoto,T., Kamahara,M. and Saito,S.
  TITLE     Novel G protein-conjugated receptor protein
  JOURNAL   Patent: JP 2001029083-A 4 06-FEB-2001;
            YAMANOUCHI PHARMACEUT CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001029083-A/4
            PD   06-FEB-2001
            PF   23-JUL-1999 JP 1999209918
            PR
            PI   ATSUSHI TAKASAKI,MITSUYUKI MATSUMOTO,TAKASHI SUGIMOTO, PI
            MASAZUMI KAMAHARA,
            PI   SATOSHI SAITO
            PC   C12N15/09,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61P25/04,
            PC   A61P25/16,
            PC   A61P25/18,C07K14/705,C12N5/10,C12P21/02,C12P21/08,C12Q1/68, PC
              G01N33/15,
            PC   G01N33/50,G01N33/53,G01N33/566,C12N15/00,A61K37/02,C12N5/00 CC

            FH   Key             Location/Qualifiers
            FT   source          1. .40
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       10 a      8 c      4 g     18 t
ORIGIN


  Query Match             51.7%;  Score 15;  DB 6;  Length 40;
  Best Local Similarity   47.8%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 uaugauucuuuuuguaagcccua 23
           :|:||::|:: : | ||| || |
Db      35 TATGATTCTTATAGAAAGTCCAA 13


RESULT   5
E50836/c
LOCUS       E50836                   40 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Novel G protein-coupled receptor.
ACCESSION   E50836
VERSION     E50836.1  GI:18633541
KEYWORDS    JP 2001054389-A/4.
SOURCE      Homo sapiens.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Takasaki,A., Matsumoto,M., Sugimoto,T., Kamahara,M. and Saito,S.
  TITLE     Novel G protein-coupled receptor
  JOURNAL   Patent: JP 2001054389-A 4 27-FEB-2001;
            YAMANOUCHI PHARMACEUT CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001054389-A/4
            PD   27-FEB-2001
            PF   17-AUG-1999 JP 1999230777
            PR
            PI   ATSUSHI TAKASAKI,MITSUYUKI MATSUMOTO,TAKASHI SUGIMOTO, PI
            MASAZUMI KAMAHARA,
            PI   SATOSHI SAITO
            PC   C12N15/09,C07K14/705,C07K16/28,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,G01N33/15,G01N33/50//C12P21/08,(C12P21/02,C12R1:91),
            PC   C12N15/00,
            PC   C12N5/00
            CC
            FH   Key             Location/Qualifiers
            FT   source          1. .40
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       10 a      8 c      4 g     18 t
ORIGIN


  Query Match           51.7%;  Score 15;  DB 6;  Length 40;
  Best Local Similarity  47.8%;  Pred. No. 2.2e+04;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccua 23
          :|:||::|:: : | ||| || |
Db     35 TATGATTCTTATAGAAAGTCCAA 13


RESULT   6
DDIDDKD/c
LOCUS       DDIDDKD                  87 bp    DNA     linear   INV 27-APR-1993
DEFINITION  D.discoideum protein kinase 4 gene, partial cds.
ACCESSION   M59747
VERSION     M59747.1  GI:167723
KEYWORDS    protein kinase 4.
SOURCE      Dictyostelium discoideum (strain AX-3) DNA.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 87)
  AUTHORS   Haribabu,B. and Dottin,R.P.
  TITLE     Identification of a protein kinase multigene family of
            Dictyostelium discoideum: Molecular cloning and expression of a
            cDNA encoding a developmentally regulated protein kinase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 1115-1119 (1991)
  MEDLINE   91142122
FEATURES             Location/Qualifiers
     source          1. .87
                     /organism="Dictyostelium discoideum"
                     /strain="AX-3"
                     /db_xref="taxon:44689"
     CDS             <1. .>87
                     /codon_start=1
                     /product="protein kinase 4"
                     /protein_id="AAA33189.1"
                     /db_xref="GI:167724"
                     /translation="NLLIDQYGHIKLTDFGFAKRITENTKSMC"
BASE COUNT       36 a     12 c     14 g     25 t
ORIGIN


  Query Match           51.7%;  Score 15;  DB 3;  Length 87;
  Best Local Similarity  39.1%;  Pred. No. 2.1e+04;
  Matches    9;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccua 23
          : :||::|:::::| ||  || |
Db     66 TGTGATTCTTTTTGCAAATCCAA 44


RESULT   7
I08597/c
LOCUS       I08597                   35 bp            linear   PAT 02-DEC-1994
DEFINITION  Sequence 12 from Patent WO 8707144.
ACCESSION   I08597
VERSION     I08597.1  GI:588701
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   Kaufman,R.J., Pittman,D.D. and Toole,J.J.J.
  TITLE     NOVEL PROCOAGULANT PROTEINS
  JOURNAL   Patent: WO 8707144-A 12 03-DEC-1987;
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="unknown"
BASE COUNT       18 a      8 c      5 g      4 t
ORIGIN


  Query Match           51.0%;  Score 14.8;  DB 6;  Length 35;
  Best Local Similarity  42.3%;  Pred. No. 2.8e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagcccuaggggcu 29
          | ::  .:::: |||| : |||||:
Db     35 GTTTTCTTTTGAAAGCTTTTGGGGCT 10


RESULT   8
AR142180/c
LOCUS       AR142180                 36 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6174696.
ACCESSION   AR142180
VERSION     AR142180.1  GI:15102480
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 36)
  AUTHORS   Seman,L.
  TITLE     Method for the determination of homocysteine
  JOURNAL   Patent: US 6174696-A 1 16-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .36
                     /organism="unknown"
BASE COUNT       12 a      7 c      9 g      8 t
ORIGIN


  Query Match           51.0%;  Score 14.8;  DB 6;  Length 36;
  Best Local Similarity  38.5%;  Pred. No. 2.8e+04;
  Matches   10;  Conservative    9;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggg 26
          :|: |  |:::::|:  ||| :| ||
Db     33 TATCAAGCTTTTTGTCCGCCATATGG 8


RESULT   9
AX115517
LOCUS       AX115517                 51 bp    DNA     linear   PAT 11-MAY-2001
```

```
DEFINITION  Sequence 640 from Patent WO0129262.
ACCESSION   AX115517
VERSION     AX115517.1  GI:14032459
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Picoult-Newburg,L. and Pohl,M.
  TITLE     Genotyping reagents, kits and methods of use thereof
  JOURNAL   Patent: WO 0129262-A 640 26-APR-2001;
            Orchid BioSciences, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       10 a      8 c      8 g     25 t
ORIGIN


  Query Match             50.3%;  Score 14.6;  DB 6;  Length 51;
  Best Local Similarity   31.0%;  Pred. No. 3.4e+04;
  Matches    9;  Conservative   11;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggggcu 29
          :|: |:: ::::: :|   |  || |||:
Db      9 TATTATTTTTTTTATAGCTCACAGTGGCT 37


RESULT  10
AF328713
LOCUS       AF328713                 51 bp    DNA     linear   ROD 29-JUN-2001
DEFINITION  Mus musculus isolate 1.2-11 T-cell receptor beta chain VDJ
            junctional region gene, partial cds.
ACCESSION   AF328713
VERSION     AF328713.1  GI:13898588
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Maryanski,J.L., Attuil,V., Hamrouni,A., Mutin,M., Rossi,M.,
            Aublin,A. and Bucher,P.
  TITLE     Individuality of Ag-selected and preimmune TCR repertoires
  JOURNAL   Immunol. Res. 23 (1), 75-84 (2001)
  MEDLINE   21310440
   PUBMED   11417861
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Maryanski,J.L., Attuil,V., Hamrouni,A., Mutin,M., Rossi,M.,
            Aublin,A. and Bucher,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-2000) INSERM Unit 503, CERVI, 21 Avenue Tony
            Garnier, 69365 Lyon Cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Mus musculus"
                     /strain="DBA/2"
                     /isolate="1.2-11"
                     /db_xref="taxon:10090"
                     /rearranged
     mRNA            <1. .>51
                     /product="T-cell receptor beta chain VDJ junctional
                     region"
     CDS             <1. .>51
                     /note="TCRBV10"
                     /codon_start=1
                     /product="T-cell receptor beta chain VDJ junctional
                     region"
                     /protein_id="AAK48813.1"
                     /db_xref="GI:13898589"
                     /translation="SAVYLCASSRGSDYTFG"
BASE COUNT        7 a     16 c     14 g     14 t
ORIGIN


  Query Match             50.3%;  Score 14.6;  DB 10;  Length 51;
  Best Local Similarity   41.4%;  Pred. No. 3.4e+04;
  Matches   12;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggggcu 29
          : :|  :|: : :|  |||  :||||||:
Db      6 TGTGTATCTCTGTGCCAGCAGTAGGGGCT 34


RESULT  11
S63972
LOCUS       S63972                   63 bp    DNA     linear   PRI 15-OCT-1993
DEFINITION  IgH (CDR3 region, VHDJH rearrangement, clone DV3-R2) [human,
            leukemia patient D, bone marrow B-cells, Genomic, 63 nt].
ACCESSION   S63972
VERSION     S63972.1  GI:408181
KEYWORDS    .
SOURCE      human leukemia patient D bone marrow B-cells.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 63)
  AUTHORS   Steenbergen,E.J., Verhagen,O.J., van Leeuwen,E.F., von dem
            Borne,A.E. and van der Schoot,C.E.
  TITLE     Distinct ongoing Ig heavy chain rearrangement processes in
            childhood B-precursor acute lymphoblastic leukemia
  JOURNAL   Blood 82 (2), 581-589 (1993)
  MEDLINE   93320440
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 135423] from the original journal article.
            This sequence comes from Fig. 3D.
FEATURES             Location/Qualifiers
     source          1. .63
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .63
                     /partial
                     /gene="IgH"
BASE COUNT       16 a     11 c     19 g     17 t
ORIGIN


  Query Match             50.3%;  Score 14.6;  DB 9;  Length 63;
  Best Local Similarity   41.4%;  Pred. No. 3.4e+04;
  Matches   12;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggggcu 29
          :|:||:  :  : |: |   |:||||||:
Db     17 TATGATAGTAGTGGTTATTACTAGGGGCT 45


RESULT  12
ECOPKYD2/c
LOCUS       ECOPKYD2                 72 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  E.coli/IS1 left junction, plasmid pKY159-TN8.
ACCESSION   M30729
VERSION     M30729.1  GI:147287
KEYWORDS    direct repeat; insertion sequence.
SEGMENT     2 of 2
SOURCE      E.coli plasmid pK159-TN8 DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 72)
  AUTHORS   Kanazawa,H., Kiyasu,T., Noumi,T., Futai,M. and Yamaguchi,K.
  TITLE     Insertions of transposable elements in the promoter proximal region
            of the gene cluster for Escherichia coli H+-ATPase: 8 base pair
            repeat generated by insertion of IS1
  JOURNAL   Mol. Gen. Genet. 194, 179-187 (1984)
```

```
  MEDLINE   84218838
FEATURES             Location/Qualifiers
     source          1. .72
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     repeat_region   39. .47
                     /note="insertion target sequence"
BASE COUNT       26 a     18 c     11 g     17 t
ORIGIN      About 720 bp after segment 1.


  Query Match             50.3%;  Score 14.6;  DB 1;  Length 72;
  Best Local Similarity   37.9%;  Pred. No. 3.3e+04;
  Matches   11;  Conservative    9;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 uaugauucuuuuuguaagcccuaggggcu 29
          :|: |  |::: : :||| | :||| | :
Db     61 TATAAAACTTTCTTTAAGGCTTAGGTGAT 33


RESULT  13
AX042583
LOCUS       AX042583                 25 bp    DNA     linear   PAT 23-NOV-2000
DEFINITION  Sequence 149 from Patent WO0065088.
ACCESSION   AX042583
VERSION     AX042583.1  GI:11341191
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Ulfendahl,P.J. and Wong,K.C.
  TITLE     Primers for identifying typing or classifying nucleic acids
  JOURNAL   Patent: WO 0065088-A 149 02-NOV-2000;
            Amersham Pharmacia Biotech AB (SE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="DQA1 Homozygote primer sequence"
BASE COUNT        3 a      3 c      5 g     14 t
ORIGIN


  Query Match             49.7%;  Score 14.4;  DB 6;  Length 25;
  Best Local Similarity   33.3%;  Pred. No. 4.3e+04;
  Matches    8;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuaggggcu 29
          :: ::::: : ||| |: ||| |:
Db      2 TTTTTTTTTTAGCACTGGGGACT 25


RESULT  14
AX043280
LOCUS       AX043280                 25 bp    DNA     linear   PAT 23-NOV-2000
DEFINITION  Sequence 846 from Patent WO0065088.
ACCESSION   AX043280
VERSION     AX043280.1  GI:11341888
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Ulfendahl,P.J. and Wong,K.C.
  TITLE     Primers for identifying typing or classifying nucleic acids
  JOURNAL   Patent: WO 0065088-A 846 02-NOV-2000;
            Amersham Pharmacia Biotech AB (SE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="DQA1 Heterozygote Primer Sequence"
BASE COUNT        3 a      3 c      5 g     14 t
ORIGIN


  Query Match             49.7%;  Score 14.4;  DB 6;  Length 25;
  Best Local Similarity   33.3%;  Pred. No. 4.3e+04;
  Matches    8;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagcccuaggggcu 29
          :: ::::: : ||| |: ||| |:
Db      2 TTTTTTTTTTAGCACTGGGGACT 25


RESULT  15
AX298174/c
LOCUS       AX298174                 35 bp    DNA     linear   PAT 26-NOV-2001
DEFINITION  Sequence 20 from Patent WO0183788.
ACCESSION   AX298174
VERSION     AX298174.1  GI:17128241
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Heitz,T., Dhondt,S., Geoffroy,P., Legrand,M. and Gouzerh,G.
  TITLE     Plant pla2 polypeptides involved in plant defence reaction,
            polyncleotides encoding said polypeptides and transformed plants
            containing them
  JOURNAL   Patent: WO 0183788-A 20 08-NOV-2001;
            Rhobio (FR)
FEATURES             Location/Qualifiers
     source          1. .35
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Oligonucleotide amorce NtPat3pGEX sens"
BASE COUNT       14 a      5 c     10 g      6 t
ORIGIN


  Query Match             49.0%;  Score 14.2;  DB 6;  Length 35;
  Best Local Similarity   40.7%;  Pred. No. 5.3e+04;
  Matches   11;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      2 augauucuuuuuguaagcccuaggggc 28
          |:  ::| ::: |:|| | |:|| | |
Db     27 ATCTTTCCTTTGGTAACCTCTAGAGTC 1


Search completed: September 13, 2002, 13:18:15
Job time: 4930 sec
```

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 13:23:15 ; Search time 280.51 Seconds
(without alignments)
177.500 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 2046006

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 29 | 100.0 | 29 | 21 | AAA70829 | Molecular interact |
| 2 | 29 | 100.0 | 29 | 21 | AAA70830 | Molecular interact |
| 3 | 29 | 100.0 | 42 | 21 | AAA71115 | Molecular interact |
| 4 | 29 | 100.0 | 42 | 21 | AAA71116 | Molecular interact |
| 5 | 29 | 100.0 | 42 | 21 | AAA71120 | Molecular interact |
| 6 | 29 | 100.0 | 42 | 21 | AAA71121 | Molecular interact |
| 7 | 29 | 100.0 | 42 | 21 | AAA71128 | Molecular interact |
| 8 | 29 | 100.0 | 42 | 21 | AAA71129 | Molecular interact |
| 9 | 28 | 96.6 | 45 | 21 | AAA70825 | Molecular interact |
| 10 | 28 | 96.6 | 45 | 21 | AAA70826 | Molecular interact |
| 11 | 28 | 96.6 | 46 | 21 | AAA71088 | Molecular interact |
| 12 | 28 | 96.6 | 46 | 21 | AAA71089 | Molecular interact |
| 13 | 28 | 96.6 | 46 | 21 | AAA71090 | Molecular interact |
| 14 | 28 | 96.6 | 46 | 21 | AAA71105 | Molecular interact |
| 15 | 28 | 96.6 | 46 | 21 | AAA71106 | Molecular interact |
| 16 | 28 | 96.6 | 46 | 21 | AAA71107 | Molecular interact |
| 17 | 24.8 | 85.5 | 42 | 21 | AAA71113 | Molecular interact |
| 18 | 24.8 | 85.5 | 42 | 21 | AAA71118 | Molecular interact |
| 19 | 24.8 | 85.5 | 42 | 21 | AAA71126 | Molecular interact |
| 20 | 23.8 | 82.1 | 46 | 21 | AAA71085 | Molecular interact |
| 21 | 23.8 | 82.1 | 46 | 21 | AAA71103 | Molecular interact |
| 22 | 23.2 | 80.0 | 29 | 21 | AAA70828 | Molecular interact |
| 23 | 23.2 | 80.0 | 42 | 21 | AAA71123 | Molecular interact |
| 24 | 23.2 | 80.0 | 42 | 21 | AAA71131 | Molecular interact |
| 25 | 22.2 | 76.6 | 45 | 21 | AAA70824 | Molecular interact |
| 26 | 22.2 | 76.6 | 46 | 21 | AAA71087 | Molecular interact |
| 27 | 22.2 | 76.6 | 46 | 21 | AAA71096 | Molecular interact |
| 28 | 22.2 | 76.6 | 46 | 21 | AAA71099 | Molecular interact |
| 29 | 22.2 | 76.6 | 46 | 21 | AAA71100 | Molecular interact |
| 30 | 22.2 | 76.6 | 46 | 21 | AAA71104 | Molecular interact |
| 31 | 21.2 | 73.1 | 42 | 21 | AAA71114 | Molecular interact |
| 32 | 21.2 | 73.1 | 42 | 21 | AAA71119 | Molecular interact |
| 33 | 21.2 | 73.1 | 42 | 21 | AAA71127 | Molecular interact |
| 34 | 21.2 | 73.1 | 46 | 21 | AAA71094 | Molecular interact |
| 35 | 21.2 | 73.1 | 46 | 21 | AAA71110 | Molecular interact |
| 36 | 20 | 69.0 | 46 | 21 | AAA71084 | Molecular interact |
| 37 | 20 | 69.0 | 46 | 21 | AAA71098 | Molecular interact |
| 38 | 20 | 69.0 | 46 | 21 | AAA71102 | Molecular interact |
| 39 | 19.6 | 67.6 | 42 | 21 | AAA71124 | Molecular interact |
| 40 | 19.6 | 67.6 | 42 | 21 | AAA71132 | Molecular interact |
| 41 | 18.6 | 64.1 | 46 | 21 | AAA71093 | Molecular interact |
| 42 | 18.6 | 64.1 | 46 | 21 | AAA71095 | Molecular interact |
| 43 | 18.6 | 64.1 | 46 | 21 | AAA71109 | Molecular interact |
| 44 | 18.6 | 64.1 | 46 | 21 | AAA71111 | Molecular interact |
| 45 | 18.4 | 63.4 | 42 | 21 | AAA71117 | Molecular interact |

ALIGNMENTS

RESULT 1
AAA70829
ID AAA70829 standard; RNA; 29 BP.
XX
AC AAA70829;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #29.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Mus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 8 A; 6 C; 6 G; 9 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||||||||||||||||||||||||||
Db 1 aaagauucuuuuuguaagccccaagggcu 29

RESULT 2
AAA70830
ID AAA70830 standard; RNA; 29 BP.
XX
AC AAA70830;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #30.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Rattus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Claim 235; Page 235; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 29 BP; 8 A; 6 C; 6 G; 9 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||||||||||||||||||||||||||
Db 1 aaagauucuuuuuguaagccccaagggcu 29

RESULT 3
AAA71115
ID AAA71115 standard; RNA; 42 BP.
XX
AC AAA71115;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #191.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||||||||||||||||||||||||||
Db 4 aaagauucuuuuuguaagccccaagggcu 32

RESULT 4
AAA71116
ID AAA71116 standard; RNA; 42 BP.
XX
AC AAA71116;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #192.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||||||||||||||||||||||||||
Db 4 aaagauucuuuuuguaagccccaagggcu 32

RESULT 5
AAA71120
ID AAA71120 standard; DNA; 42 BP.
XX
AC AAA71120;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #126.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 T; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 69.0%; Pred. No. 0.001;
Matches 20; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||:|:::::|:|||||||||||||:
Db 4 aaagattcttttttgtaagccccaagggct 32

RESULT 6
AAA71121
ID AAA71121 standard; DNA; 42 BP.
XX
AC AAA71121;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #127.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 125; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 T; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 69.0%; Pred. No. 0.001;
Matches 20; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||:|:::::|:|||||||||||||:
Db 4 aaagattcttttttgtaagccccaagggct 32

RESULT 7
AAA71128
ID AAA71128 standard; RNA; 42 BP.
XX
AC AAA71128;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #197.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||||||||||||||||||||||||||
Db 4 aaagauucuuuuuguaagccccaagggcu 32

RESULT 8
AAA71129
ID AAA71129 standard; RNA; 42 BP.
XX
AC AAA71129;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #198.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 126; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 42 BP; 13 A; 7 C; 7 G; 15 U; 0 other;

Query Match 100.0%; Score 29; DB 21; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggcu 29
||||||||||||||||||||||||||||||
Db 4 aaagauucuuuuuguaagccccaagggcu 32

RESULT 9
AAA70825
ID AAA70825 standard; RNA; 45 BP.
XX
AC AAA70825;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #25.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Mus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Claim 221; Page 232; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 45 BP; 14 A; 7 C; 9 G; 15 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggc 28
||||||||||||||||||||||||||||
Db 18 aaagauucuuuuuguaagccccaagggc 45

RESULT 10
AAA70826
ID AAA70826 standard; RNA; 45 BP.
XX
AC AAA70826;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #26.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Rattus sp.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Claim 222; Page 232; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 45 BP; 14 A; 7 C; 9 G; 15 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggc 28
||||||||||||||||||||||||||||
Db 18 aaagauucuuuuuguaagccccaagggc 45

RESULT 11
AAA71088
ID AAA71088 standard; DNA; 46 BP.
XX
AC AAA71088;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #111.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 71.4%; Pred. No. 0.0029;
Matches 20; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggc 28
||||||::|::::|:||||||||||||||
Db 19 aaagattcttttgtaagccccaagggc 46

RESULT 12
AAA71089
ID AAA71089 standard; DNA; 46 BP.
XX
AC AAA71089;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #112.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 71.4%; Pred. No. 0.0029;
Matches 20; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggc 28
||||||::|::::|:||||||||||||||
Db 19 aaagattcttttgtaagccccaagggc 46

RESULT 13
AAA71090
ID AAA71090 standard; DNA; 46 BP.
XX
AC AAA71090;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site DNA #113.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 121; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 T; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 71.4%; Pred. No. 0.0029;
Matches 20; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggc 28
||||||::|:::::|:||||||||||||||
Db 19 aaagattctttttgtaagccccaagggc 46

RESULT 14
AAA71105
ID AAA71105 standard; RNA; 46 BP.
XX
AC AAA71105;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #181.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggc 28
||||||||||||||||||||||||||||
Db 19 aaagauucuuuuuguaagccccaagggc 46

RESULT 15
AAA71106
ID AAA71106 standard; RNA; 46 BP.
XX
AC AAA71106;
XX
DT 27-APR-2001 (first entry)
XX
DE Molecular interaction site RNA #182.
XX
KW Modulator; identification; molecular interaction; virtual library; ss.
XX
OS Unidentified.
XX
PN WO9958947-A2.
XX
PD 18-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10361.
XX
PR 12-MAY-1998; 98US-0076404.
PR 12-MAY-1998; 98US-0085092.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ecker DJ, Griffey R, Crooke ST, Sampath R, Swayze E, Mohan V;
PI Hofstadler S, McNeil J;
XX
DR WPI; 2000-086439/07.
XX
PT Identifying compounds which modulate activity of target biomolecules,
PT used to provide compounds which can be used as pharmacological,

PT agricultural and industrial compounds -
XX
PS Example 7; Figure 122; 405pp; English.
XX
CC This invention describes a novel method for identifying compounds which
CC modulate the activity of a target biomolecule. The method uses
CC 3-dimensional representations of the biomolecule and a library of
CC compounds and comprises (a) identifying at least one molecular
CC interaction site of the target RNA; (b) generating in silico a virtual
CC library of compounds predicted or calculated to interact with the
CC molecular interaction site; and (c) comparing 3-dimensional (3-D)
CC representations of the target RNA with members of the virtual library of
CC compounds to generate a hierarchy of the compounds ranked in accordance
CC with their respective ability to form physical interactions with the
CC molecular interaction site. The method also describes (1) RNA comprising
CC a joined sequence of at least 24 nucleotides but not more than 70
CC nucleotides and having secondary structure defined by: (a) 3 nucleotides
CC forming a first side of a first double stranded (ds) region; (b) 2
CC nucleotides forming a first side of an internal loop region; (c) 4
CC nucleotides forming a first side of a second ds region; (d) 4 or 5
CC nucleotides forming an end loop region; (e) 4 nucleotides forming a
CC second side of the second ds region; (f) 4 nucleotides forming a second
CC side of the internal loop region; and (g) 3 nucleotides forming a second
CC side of the first ds region; (2) a purified and isolated RNA fragment
CC comprising the human sequence UUUACAACAUAAUCUAGUUUACAGAAAAAUC (II). The
CC methods and products can be used for identifying agents which modulate
CC the activity of biomolecules, particularly RNA. Such agents can be used
CC as pharmaceutical, agricultural or industrial compounds.
XX
SQ Sequence 46 BP; 14 A; 7 C; 9 G; 16 U; 0 other;

Query Match 96.6%; Score 28; DB 21; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagccccaagggc 28
||||||||||||||||||||||||||||
Db 19 aaagauucuuuuuguaagccccaagggc 46

Search completed: September 13, 2002, 13:23:15
Job time: 5110 sec

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 13:18:15 ; Search time 2058.64 Seconds
(without alignments)
294.791 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 843946

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| c 1 | 16.2 | 55.9 | 48 | 6 | AX018731 | AX018731 Sequence |
| c 2 | 15.4 | 53.1 | 73 | 6 | E02131 | E02131 Pseudoknot |
| c 3 | 15.2 | 52.4 | 33 | 6 | AR020509 | AR020509 Sequence |
| 4 | 15 | 51.7 | 23 | 23 | E09974 | E09974 Primer for |
| 5 | 15 | 51.7 | 23 | 23 | E10118 | E10118 PCR primer |
| c 6 | 14.8 | 51.0 | 29 | 6 | AR019319 | AR019319 Sequence |
| c 7 | 14.8 | 51.0 | 29 | 6 | AR061847 | AR061847 Sequence |
| c 8 | 14.8 | 51.0 | 29 | 6 | AR147578 | AR147578 Sequence |
| c 9 | 14.8 | 51.0 | 29 | 6 | I34733 | I34733 Sequence 25 |
| c 10 | 14.8 | 51.0 | 29 | 6 | I67987 | I67987 Sequence 25 |
| c 11 | 14.8 | 51.0 | 32 | 6 | AR061867 | AR061867 Sequence |
| 12 | 14.8 | 51.0 | 80 | 6 | A52215 | A52215 Sequence 5 |
| 13 | 14.6 | 50.3 | 25 | 6 | AX043294 | AX043294 Sequence |
| c 14 | 14.6 | 50.3 | 53 | 6 | AR061021 | AR061021 Sequence |
| c 15 | 14.6 | 50.3 | 87 | 3 | DDIDDKD | M59747 D.discoideu |
| 16 | 14.4 | 49.7 | 25 | 6 | AX043671 | AX043671 Sequence |
| 17 | 14.4 | 49.7 | 51 | 6 | AX117185 | AX117185 Sequence |
| 18 | 14.4 | 49.7 | 81 | 14 | AF166557 | AF166557 Hepatitis |
| 19 | 14.4 | 49.7 | 81 | 14 | AF166559 | AF166559 Hepatitis |
| 20 | 14.4 | 49.7 | 81 | 14 | AF166560 | AF166560 Hepatitis |
| 21 | 14.4 | 49.7 | 81 | 14 | AF166561 | AF166561 Hepatitis |
| 22 | 14.2 | 49.0 | 69 | 6 | AR052906 | AR052906 Sequence |
| 23 | 14.2 | 49.0 | 69 | 6 | AR054269 | AR054269 Sequence |
| 24 | 14.2 | 49.0 | 69 | 6 | AR054471 | AR054471 Sequence |
| c 25 | 14.2 | 49.0 | 87 | 9 | HSLASSBIN | X96992 H.sapiens g |
| c 26 | 14.2 | 49.0 | 100 | 5 | AF174511 | AF174511 Bufo mela |
| c 27 | 14.2 | 49.0 | 100 | 5 | AF174512 | AF174512 Bufo mela |
| c 28 | 14.2 | 49.0 | 100 | 5 | AF174513 | AF174513 Bufo mela |
| c 29 | 14.2 | 49.0 | 100 | 5 | AF174514 | AF174514 Bufo mela |
| c 30 | 14.2 | 49.0 | 100 | 5 | AF174515 | AF174515 Bufo mela |
| c 31 | 14.2 | 49.0 | 100 | 5 | AF174516 | AF174516 Bufo mela |
| c 32 | 14.2 | 49.0 | 100 | 5 | AF174517 | AF174517 Bufo mela |
| c 33 | 14.2 | 49.0 | 100 | 5 | AF174518 | AF174518 Bufo mela |
| c 34 | 14.2 | 49.0 | 100 | 5 | AF174519 | AF174519 Bufo mela |
| c 35 | 14.2 | 49.0 | 100 | 5 | AF174520 | AF174520 Bufo mela |
| c 36 | 14.2 | 49.0 | 100 | 5 | AF174521 | AF174521 Bufo mela |
| 37 | 14.2 | 49.0 | 100 | 9 | HUM0401I | M83920 Human MHC c |
| 38 | 14.2 | 49.0 | 100 | 9 | HUM0501I | M83921 Human MHC c |
| 39 | 14 | 48.3 | 25 | 6 | AX043055 | AX043055 Sequence |
| 40 | 14 | 48.3 | 41 | 6 | AX316543 | AX316543 Sequence |
| c 41 | 14 | 48.3 | 75 | 9 | AF312283 | AF312283 Homo sapi |
| 42 | 14 | 48.3 | 100 | 10 | RNU12531 | U12531 Rattus norv |
| c 43 | 13.8 | 47.6 | 27 | 6 | AX249915 | AX249915 Sequence |
| c 44 | 13.8 | 47.6 | 38 | 9 | HSTCRDV21 | X69283 H.sapiens m |
| c 45 | 13.8 | 47.6 | 38 | 9 | HSTCRDV25 | X69287 H.sapiens m |

ALIGNMENTS

RESULT 1
AX018731/c
LOCUS AX018731 48 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 20 from Patent WO9944633.
ACCESSION AX018731
VERSION AX018731.1 GI:10042853
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 48)
AUTHORS Minke,J.M. and Audonnet,J.C.
TITLE Live recombined vaccines injected with adjuvant
JOURNAL Patent: WO 9944633-A 20 10-SEP-1999;
MINKE JULES MAARTEN (FR); MERIAL SAS (FR); AUDONNET JEAN CHRISTOPHE FRANC (FR)
FEATURES Location/Qualifiers
source 1. .48
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="oligonucleotide"
BASE COUNT 17 a 9 c 9 g 13 t
ORIGIN

```
  Query Match            55.9%;  Score 16.2;  DB 6;  Length 48;
  Best Local Similarity   44.8%;  Pred. No. 5.9e+03;
  Matches   13;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 aaagauucuuuuuguaagccccaagggcu 29
          |||  :  :::::|:||||  |  ||||:
Db     37 AAATCTAATTTTTGTAAGCTTCCCGGGCT 9


RESULT   2
E02131/c
LOCUS       E02131                    73 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  Pseudoknot region of 3'non-coding region.
ACCESSION   E02131
VERSION     E02131.1  GI:5708465
KEYWORDS    JP 1989281079-A/3.
SOURCE      Tobacco mosaic virus.
  ORGANISM  Tobacco mosaic virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
REFERENCE   1  (bases 1 to 73)
  AUTHORS   Okada,Y. and Takamatsu,N.
  TITLE     ATTENUATED STRAIN OF PLANT VIRUS AND PREPARATION THEREOF
  JOURNAL   Patent: JP 1989281079-A 3 13-NOV-1989;
            KIRIN BREWERY CO LTD
COMMENT     OS   Tobacco mosaic virus
            PN   JP 1989281079-A/3
            PD   13-NOV-1989
            PF   09-MAY-1988 JP 1988110353
            PI   OKADA YOSHIMI, TAKAMATSU NOBUHIKO
            PC   C12N7/04,C12N15/00,(C12N15/00,C12R1:91);
            CC   strandedness: Single;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=CcTMV;
            FH   Key             Location/Qualifiers
            FH
            FT   3'UTR           <1. .>73.
FEATURES              Location/Qualifiers
     source           1. .73
                      /organism="Tobacco mosaic virus"
                      /db_xref="taxon:12242"
BASE COUNT       21 a      17 c      17 g      18 t
ORIGIN


  Query Match            53.1%;  Score 15.4;  DB 6;  Length 73;
  Best Local Similarity   52.0%;  Pred. No. 1.4e+04;
  Matches   13;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagggc 28
          | :: :::  |:|||||| ||| |
Db     43 GCTTATTTCGGTAAGCCCCTAGGTC 19


RESULT   3
AR020509/c
LOCUS       AR020509                  33 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 5 from patent US 5789171.
ACCESSION   AR020509
VERSION     AR020509.1  GI:3975124
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Smeltzer,M.S.
  TITLE     Use of cna, fnba, fnbb, and hlb, gene probes for the
            strain-specific identification of Staphylococcus aureus
  JOURNAL   Patent: US 5789171-A 5 04-AUG-1998;
FEATURES              Location/Qualifiers
     source           1. .33
                      /organism="unknown"
BASE COUNT       12 a      8 c      7 g      6 t
ORIGIN


  Query Match            52.4%;  Score 15.2;  DB 6;  Length 33;
  Best Local Similarity   42.9%;  Pred. No. 1.8e+04;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      2 aagauucuuuuuguaagccccaagggcu 29
          | ||:: :::: |:||   ||  ||||:
Db     32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT   4
E09974
ID   E09974     standard; DNA; UNC; 23 BP.
XX
AC   E09974;
XX
SV   E09974.1
XX
DT   07-OCT-1997 (Rel. 52, Created)
DT   02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE   Primer for amplifying human herpes virus.
XX
KW   JP 1995250699-A/20.
XX
OS   unidentified
OC   unclassified.
XX
RN   [1]
RP   1-23
RA   Yamanishi K., Mukai T., Aono T., Kondo M., Takarada Y.;
RT   "METHOD FOR DISCRIMINATORY DETECTION OF HUMAN HERPES VIRUS AND REAGENT
RT   THEREFOR";
RL   Patent number JP1995250699-A/20, 03-OCT-1995.
RL   TOYOBO CO LTD.
XX
CC   OS   None
CC   OC   Artificial sequences.
CC   PN   JP 1995250699-A/20
CC   PD   03-OCT-1995
CC   PF   11-MAR-1994 JP 1994041101
CC   PI   YAMANISHI KOICHI, MUKAI TORU, AONO TOSHIYA, KONDO MOTOHIRO,
CC   PI   TAKARADA YUTAKA
CC   PC   C12Q1/68,C12N15/09,C12Q1/70;
CC   CC   strandedness: Single;
CC   CC   topology: Linear;
CC   CC   hypothetical: No;
CC   FH   Key             Location/Qualifiers
CC   FH
CC   FT   source          1. .23
CC   FT                   /organism="Artificial sequences"
CC   FT   misc_feature    1. .23
CC   FT                   /note="Common sequences for human herpes virus
CC   FT                   6-type A,
CC   FT                   human herpes virus 6-type B,human herpes
CC   FT                   virus-7 and
CC   FT                   cytomegalovirus"
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .23
FT                   /db_xref="taxon:32644"
FT                   /organism="unidentified"
XX
SQ   Sequence 23 BP; 5 A; 5 C; 6 G; 7 T; 0 other;


  Query Match            51.7%;  Score 15;  DB 23;  Length 23;
```

```
  Best Local Similarity   52.2%;  Pred. No. 2.2e+04;
  Matches   12;  Conservative    6;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagg 26
          ||: |:::::|  |  |||||||
Db      1 GATCCTTTTTGAGATGCCCAAGG 23


RESULT   5
E10118
ID   E10118     standard; DNA; UNC; 23 BP.
XX
AC   E10118;
XX
SV   E10118.1
XX
DT   08-OCT-1997 (Rel. 52, Created)
DT   02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE   PCR primer to detect human herpes virus-6A,6B and 7.
XX
KW   JP 1995284391-A/6.
XX
OS   unidentified
OC   unclassified.
XX
RN   [1]
RP   1-23
RA   Yamanishi K., Mukai T., Aono T., Kondo M., Takarada Y.;
RT   "OLIGONUCLEOTIDE FOR DETECTING HUMAN HERPES VIRUS AND ITS USE";
RL   Patent number JP1995284391-A/6, 31-OCT-1995.
RL   TOYOBO CO LTD.
XX
CC   OS   None
CC   OC   Artificial sequences.
CC   PN   JP 1995284391-A/6
CC   PD   31-OCT-1995
CC   PF   19-APR-1994 JP 1994080488
CC   PI   YAMANISHI KOICHI, MUKAI TORU, AONO TOSHIYA, KONDO MOTOHIRO,
CC   PI   TAKARADA YUTAKA
CC   PC   C12N15/09,C07H21/04,C12Q1/68,G01N33/567,G01N33/569;
CC   CC   strandedness: Single;
CC   CC   topology: Linear;
CC   FH   Key             Location/Qualifiers
CC   FH
CC   FT   source          1. .23
CC   FT                   /organism="Artificial sequences"
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .23
FT                   /db_xref="taxon:32644"
FT                   /organism="unidentified"
XX
SQ   Sequence 23 BP; 5 A; 5 C; 6 G; 7 T; 0 other;


  Query Match             51.7%;  Score 15;  DB 23;  Length 23;
  Best Local Similarity   52.2%;  Pred. No. 2.2e+04;
  Matches   12;  Conservative    6;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagg 26
          ||: |:::::|  |  |||||||
Db      1 GATCCTTTTTGAGATGCCCAAGG 23


RESULT   6
AR019319/c
LOCUS       AR019319                 29 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 25 from patent US 5783406.
ACCESSION   AR019319
VERSION     AR019319.1  GI:3974433
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W. and Puers,C.
  TITLE     Allelic ladders for short tandem repeat loci
  JOURNAL   Patent: US 5783406-A 25 21-JUL-1998;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagggcu 29
          ||:: :: :: : | || | |||||:
Db     29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   7
AR061847/c
LOCUS       AR061847                 29 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 39 from patent US 5843660.
ACCESSION   AR061847
VERSION     AR061847.1  GI:5989538
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W., Micka,K.A. and Rabbach,D.R.
  TITLE     Multiplex amplification of short tandem repeat loci
  JOURNAL   Patent: US 5843660-A 39 01-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagggcu 29
          ||:: :: :: : | || | |||||:
Db     29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   8
AR147578/c
LOCUS       AR147578                 29 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 31 from patent US 6221598.
ACCESSION   AR147578
VERSION     AR147578.1  GI:15111381
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W., Sprecher,C.J. and Lins,A.M.
  TITLE     Multiplex amplification of short tandem repeat loci
  JOURNAL   Patent: US 6221598-A 31 24-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN
```

```
  Query Match             51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   9
I34733/c
LOCUS       I34733                   29 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 25 from patent US 5599666.
ACCESSION   I34733
VERSION     I34733.1  GI:2087701
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
  REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W. and Puers,C.
  TITLE     Allelic ladders for short tandem repeat loci
  JOURNAL   Patent: US 5599666-A 25 04-FEB-1997;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT  10
I67987/c
LOCUS       I67987                   29 bp    DNA     linear   PAT 04-FEB-1998
DEFINITION  Sequence 25 from patent US 5674686.
ACCESSION   I67987
VERSION     I67987.1  GI:2830109
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
  REFERENCE   1  (bases 1 to 29)
  AUTHORS   Schumm,J.W. and Puers,C.
  TITLE     Allelic ladders for short tandem repeat loci
  JOURNAL   Patent: US 5674686-A 25 07-OCT-1997;
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="unknown"
BASE COUNT       12 a      4 c      7 g      6 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 6;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT  11
AR061867/c
LOCUS       AR061867                 32 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 59 from patent US 5843660.
ACCESSION   AR061867
VERSION     AR061867.1  GI:5989558
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 32)
  AUTHORS   Schumm,J.W., Micka,K.A. and Rabbach,D.R.
  TITLE     Multiplex amplification of short tandem repeat loci
  JOURNAL   Patent: US 5843660-A 59 01-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .32
                     /organism="unknown"
BASE COUNT       13 a      4 c      8 g      7 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 6;  Length 32;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+04;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT  12
A52215
LOCUS       A52215                   80 bp    DNA     linear   PAT 12-DEC-1997
DEFINITION  Sequence 5 from Patent EP0705842.
ACCESSION   A52215
VERSION     A52215.1  GI:2833366
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 80)
  AUTHORS   Bartnik,E.D. and Margerie,D.D.
  TITLE     Regulated genes by stimulation of chondrocytes with 1L-1beta
  JOURNAL   Patent: EP 0705842-A 5 10-APR-1996;
            HOECHST AG (DE)
COMMENT     Other publication ZA 9508381 960424
            Other publication JP 8191693 960730
            Other publication CA 2159957 960407
            Other publication AU 3308695 960418.
FEATURES             Location/Qualifiers
     source          1. .80
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       20 a      22 c      17 g      21 t
ORIGIN


  Query Match             51.0%;  Score 14.8;  DB 6;  Length 80;
  Best Local Similarity   50.0%;  Pred. No. 2.7e+04;
  Matches   13;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 aaagauucuuuuuguaagccccaagg 26
           | | |:: :: ::| | |||||| ||
Db      13 ACATATTTTTATTGGAGGCCCCATGG 38


RESULT  13
AX043294
LOCUS       AX043294                 25 bp    DNA     linear   PAT 23-NOV-2000
DEFINITION  Sequence 860 from Patent WO0065088.
ACCESSION   AX043294
VERSION     AX043294.1  GI:11341902
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
```

```
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Ulfendahl,P.J. and Wong,K.C.
  TITLE     Primers for identifying typing or classifying nucleic acids
  JOURNAL   Patent: WO 0065088-A 860 02-NOV-2000;
            Amersham Pharmacia Biotech AB (SE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="DQA1 Heterozygote Primer Sequence"
BASE COUNT        4 a      4 c      3 g     14 t
ORIGIN


  Query Match             50.3%;  Score 14.6;  DB 6;  Length 25;
  Best Local Similarity   42.9%;  Pred. No. 3.4e+04;
  Matches    9;  Conservative    8;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagccccaagg 26
          :: ::::::: |:|||| ||| |
Db      5 TTTTTTTTGTCAGCCACAATG 25


RESULT  14
AR061021/c
LOCUS       AR061021                  53 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 46 from patent US 5843456.
ACCESSION   AR061021
VERSION     AR061021.1  GI:5988712
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 53)
  AUTHORS   Paoletti,E. and Maki,J.
  TITLE     Alvac poxvirus-rabies compositions and combination compositions and
            uses
  JOURNAL   Patent: US 5843456-A 46 01-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .53
                     /organism="unknown"
BASE COUNT       20 a     10 c     10 g     13 t
ORIGIN


  Query Match             50.3%;  Score 14.6;  DB 6;  Length 53;
  Best Local Similarity   47.6%;  Pred. No. 3.4e+04;
  Matches   10;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      9 uuuuuguaagccccaagggcu 29
          :::::|:|||| |  ||||:
Db     29 TTTTTGTAAGCTTCCCGGGCT 9


RESULT  15
DDIDDKD/c
LOCUS       DDIDDKD                   87 bp    DNA     linear   INV 27-APR-1993
DEFINITION  D.discoideum protein kinase 4 gene, partial cds.
ACCESSION   M59747
VERSION     M59747.1  GI:167723
KEYWORDS    protein kinase 4.
SOURCE      Dictyostelium discoideum (strain AX-3) DNA.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 87)
  AUTHORS   Haribabu,B. and Dottin,R.P.
  TITLE     Identification of a protein kinase multigene family of
            Dictyostelium discoideum: Molecular cloning and expression of a
            cDNA encoding a developmentally regulated protein kinase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 1115-1119 (1991)
  MEDLINE   91142122
FEATURES             Location/Qualifiers
     source          1. .87
                     /organism="Dictyostelium discoideum"
                     /strain="AX-3"
                     /db_xref="taxon:44689"
     CDS             <1. .>87
                     /codon_start=1
                     /product="protein kinase 4"
                     /protein_id="AAA33189.1"
                     /db_xref="GI:167724"
                     /translation="NLLIDQYGHIKLTDFGFAKRITENTKSMC"
BASE COUNT       36 a     12 c     14 g     25 t
ORIGIN


  Query Match             50.3%;  Score 14.6;  DB 3;  Length 87;
  Best Local Similarity   47.6%;  Pred. No. 3.4e+04;
  Matches   10;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaa 24
          ||::|::::::| ||  || ||
Db     63 GATTCTTTTTGCAAATCCAAA 43


Search completed: September 13, 2002, 13:18:18
Job time: 4933 sec
```

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 12:37:52 ; Search time 63.83 Seconds
(without alignments)
111.599 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 613726

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| c | 1 | 15.2 | 52.4 | 33 | 1 | US-08-667-079B-5 | Sequence 5, Appli |
| c | 2 | 15.2 | 52.4 | 75 | 1 | US-07-971-101-6 | Sequence 6, Appli |
| c | 3 | 14.8 | 51.0 | 29 | 1 | US-08-219-633-25 | Sequence 25, Appl |
| c | 4 | 14.8 | 51.0 | 29 | 1 | US-08-515-236-25 | Sequence 25, Appl |
| c | 5 | 14.8 | 51.0 | 29 | 1 | US-08-761-950-25 | Sequence 25, Appl |
| c | 6 | 14.8 | 51.0 | 29 | 2 | US-08-632-575B-39 | Sequence 39, Appl |
| c | 7 | 14.8 | 51.0 | 29 | 4 | US-09-327-229-31 | Sequence 31, Appl |
| c | 8 | 14.8 | 51.0 | 29 | 5 | PCT-US95-12608-31 | Sequence 31, Appl |
| c | 9 | 14.8 | 51.0 | 32 | 2 | US-08-632-575B-59 | Sequence 59, Appl |
| c | 10 | 14.6 | 50.3 | 53 | 2 | US-08-486-969-46 | Sequence 46, Appl |
| | 11 | 14.2 | 49.0 | 69 | 2 | US-08-410-654B-30 | Sequence 30, Appl |
| | 12 | 14.2 | 49.0 | 69 | 2 | US-08-474-851-30 | Sequence 30, Appl |
| | 13 | 14.2 | 49.0 | 69 | 2 | US-08-481-560-30 | Sequence 30, Appl |
| c | 14 | 13.8 | 47.6 | 25 | 4 | US-08-943-731-336 | Sequence 336, App |
| | 15 | 13.6 | 46.9 | 41 | 4 | US-09-565-156A-2 | Sequence 2, Appli |
| c | 16 | 13.6 | 46.9 | 79 | 1 | US-08-472-255A-136 | Sequence 136, App |
| c | 17 | 13.6 | 46.9 | 79 | 1 | US-08-479-724A-136 | Sequence 136, App |
| c | 18 | 13.6 | 46.9 | 79 | 3 | US-08-472-256B-136 | Sequence 136, App |
| c | 19 | 13.6 | 46.9 | 79 | 4 | US-08-952-793-136 | Sequence 136, App |
| c | 20 | 13.6 | 46.9 | 79 | 5 | PCT-US96-09455A-136 | Sequence 136, App |
| | 21 | 13.6 | 46.9 | 82 | 4 | US-09-565-156A-10 | Sequence 10, Appl |
| c | 22 | 13.6 | 46.9 | 82 | 4 | US-09-565-156A-23 | Sequence 23, Appl |
| c | 23 | 13.6 | 46.9 | 93 | 4 | US-09-565-156A-11 | Sequence 11, Appl |
| c | 24 | 13.6 | 46.9 | 93 | 4 | US-09-565-156A-13 | Sequence 13, Appl |
| c | 25 | 13.6 | 46.9 | 93 | 4 | US-09-565-156A-15 | Sequence 15, Appl |
| | 26 | 13.4 | 46.2 | 32 | 3 | US-08-718-738-16 | Sequence 16, Appl |
| | 27 | 13.4 | 46.2 | 32 | 4 | US-09-221-844-16 | Sequence 16, Appl |
| | 28 | 13.4 | 46.2 | 32 | 5 | PCT-US95-03323A-16 | Sequence 16, Appl |
| c | 29 | 13.4 | 46.2 | 46 | 1 | US-08-171-389-42 | Sequence 42, Appl |
| c | 30 | 13.4 | 46.2 | 46 | 1 | US-08-171-389-45 | Sequence 45, Appl |
| c | 31 | 13.4 | 46.2 | 46 | 1 | US-08-123-936-42 | Sequence 42, Appl |
| c | 32 | 13.4 | 46.2 | 46 | 1 | US-08-123-936-45 | Sequence 45, Appl |
| c | 33 | 13.4 | 46.2 | 46 | 2 | US-08-475-228A-42 | Sequence 42, Appl |
| c | 34 | 13.4 | 46.2 | 46 | 2 | US-08-475-228A-45 | Sequence 45, Appl |
| c | 35 | 13.4 | 46.2 | 46 | 3 | US-08-482-080A-42 | Sequence 42, Appl |
| c | 36 | 13.4 | 46.2 | 46 | 3 | US-08-482-080A-45 | Sequence 45, Appl |
| c | 37 | 13.4 | 46.2 | 46 | 5 | PCT-US93-12388-42 | Sequence 42, Appl |
| c | 38 | 13.4 | 46.2 | 46 | 5 | PCT-US93-12388-45 | Sequence 45, Appl |
| | 39 | 13.4 | 46.2 | 50 | 1 | US-08-245-754A-13 | Sequence 13, Appl |
| c | 40 | 13.4 | 46.2 | 50 | 1 | US-08-171-389-46 | Sequence 46, Appl |
| c | 41 | 13.4 | 46.2 | 50 | 1 | US-08-123-936-46 | Sequence 46, Appl |
| c | 42 | 13.4 | 46.2 | 50 | 2 | US-08-475-228A-46 | Sequence 46, Appl |
| | 43 | 13.4 | 46.2 | 50 | 2 | US-08-597-731-13 | Sequence 13, Appl |
| c | 44 | 13.4 | 46.2 | 50 | 3 | US-08-482-080A-46 | Sequence 46, Appl |
| c | 45 | 13.4 | 46.2 | 50 | 5 | PCT-US93-12388-46 | Sequence 46, Appl |

ALIGNMENTS

```
RESULT   1
US-08-667-079B-5/c
; Sequence 5, Application US/08667079B
; Patent No. 5789171
;  GENERAL INFORMATION:
;    APPLICANT:  Mark S. Smeltzer
;    TITLE OF INVENTION:  Use of cna, fnbA, fnbB, and hlb Gene Probes for the Strain-
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Benjamin Aaron Adler, MCGREGOR & ADLER, P.C.
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/667,079B
;      FILING DATE:  June 20, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler, Benjamin Aaron
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5886
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    ORIGINAL SOURCE:
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
US-08-667-079B-5
```

```
  Query Match             52.4%;  Score 15.2;  DB 1;  Length 33;
  Best Local Similarity   42.9%;  Pred. No. 1.9e+02;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 aagauucuuuuuguaagccccaagggcu 29
           | ||:: :::: |:||   ||  ||||:
Db      32 ATGATTGTTTTAGTAATTTCCCCGGGCT 5


RESULT   2
US-07-971-101-6/c
; Sequence 6, Application US/07971101
; Patent No. 5443969
;  GENERAL INFORMATION:
;    APPLICANT:  Wilson, Thomas M.A. et al.
;    TITLE OF INVENTION:  RNA Packaging System
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/971,101
;      FILING DATE:  19921029
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7108-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  75 nucleotides
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  RNA
US-07-971-101-6


  Query Match             52.4%;  Score 15.2;  DB 1;  Length 75;
  Best Local Similarity   53.6%;  Pred. No. 2.2e+02;
  Matches   15;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 aaagauucuuuuuguaagccccaagggc 28
           || ||  :  :: :|:|||  ||| ||||
Db      41 AACGACTTCTTCTGTAAGTTCCATGGGC 14


RESULT   3
US-08-219-633-25/c
; Sequence 25, Application US/08219633
; Patent No. 5599666
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Puers, Christoph
;    TITLE OF INVENTION:  ALLELIC LADDERS FOR SHORT TANDEM REPEAT
;    TITLE OF INVENTION:  LOCI
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  One South Pinckney Street, P.O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/219,633
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 257-5353
;      TELEFAX:  (608) 257-9175
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-219-633-25


  Query Match             51.0%;  Score 14.8;  DB 1;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   4
US-08-515-236-25/c
; Sequence 25, Application US/08515236
; Patent No. 5674686
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Puers, Christoph
;    TITLE OF INVENTION:  ALLELIC LADDERS FOR SHORT TANDEM REPEAT
;    TITLE OF INVENTION:  LOCI
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  One South Pinckney Street, P.O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/515,236
;      FILING DATE:  15-AUG-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/219,633
;      FILING DATE:  28-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.019
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (608) 257-5353
;       TELEFAX:  (608) 257-9175
;   INFORMATION FOR SEQ ID NO:  25:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  29 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-515-236-25


  Query Match             51.0%;  Score 14.8;  DB 1;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   5
US-08-761-950-25/c
; Sequence 25, Application US/08761950
; Patent No. 5783406
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Puers, Christoph
;    TITLE OF INVENTION:  ALLELIC LADDERS FOR SHORT TANDEM REPEAT
;    TITLE OF INVENTION:  LOCI
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  One South Pinckney Street, P.O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/761,950
;      FILING DATE:  09-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/515,236
;      FILING DATE:  15-AUG-1995
;      APPLICATION NUMBER:  US 08/219,633
;      FILING DATE:  28-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.019
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 257-5353
;      TELEFAX:  (608) 257-9175
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-761-950-25


  Query Match             51.0%;  Score 14.8;  DB 1;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   6
US-08-632-575B-39/c
; Sequence 39, Application US/08632575B
; Patent No. 5843660
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    TITLE OF INVENTION:  Multiplex Amplification of
;    TITLE OF INVENTION:  Short Tandem Repeat Loci
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Promega Corporation
;      STREET:  2800 Woods Hollow Road
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53711-5399
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Mb
;      COMPUTER:  IBM compatible PC
;      OPERATING SYSTEM:  DOS, version 6.0
;      SOFTWARE:  WordPerfect 5.1 (DOS text format)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/632,575B
;      FILING DATE:  04/15/96
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/316,544
;      FILING DATE:  09/30/94
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    POSITION IN GENOME:
;      MAP POSITION:  HUMvWFA31
US-08-632-575B-39


  Query Match             51.0%;  Score 14.8;  DB 2;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagccccaagggcu 29
           ||:: :: :: : | || | |||||:
Db      29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   7
US-09-327-229-31/c
; Sequence 31, Application US/09327229
; Patent No. 6221598
;   GENERAL INFORMATION:
;       APPLICANT: Schumm, James W.
;                  Sprecher, Cynthia J.
;                  Lins, Ann M.
;       TITLE OF INVENTION: MULTIPLEX AMPLIFICATION OF SHORT TANDEM
;                           REPEAT LOCI
;       NUMBER OF SEQUENCES: 32
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Ross & Stevens, S.C.
;            STREET: P. O. Box 2599
;            CITY: Madison
;            STATE: Wisconsin
;            COUNTRY: U.S.A.
;            ZIP: 53701-2599
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
;           SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/327,229
;           FILING DATE: 07-Jun-1999
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US/08/316,544
;           FILING DATE: 30-SEP-1994
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Sara, Charles S.
;           REGISTRATION NUMBER: 30,492
;           REFERENCE/DOCKET NUMBER: 34506.022
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 608-257-5353
;           TELEFAX: 608-257-9175
;   INFORMATION FOR SEQ ID NO: 31:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 29 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-327-229-31


  Query Match             51.0%;  Score 14.8;  DB 4;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagggcu 29
          ||:: :: :: : | || | |||||:
Db     29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   8
PCT-US95-12608-31/c
; Sequence 31, Application PC/TUS9512608
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    APPLICANT:  Sprecher, Cynthia J.
;    APPLICANT:  Lins, Ann M.
;    TITLE OF INVENTION:  MULTIPLEX AMPLIFICATION OF SHORT TANDEM
;    TITLE OF INVENTION:  REPEAT LOCI
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ross & Stevens, S.C.
;      STREET:  P. O. Box 2599
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53701-2599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/12608
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REGISTRATION NUMBER:  30,492
;      REFERENCE/DOCKET NUMBER:  34506.022
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-257-5353
;      TELEFAX:  608-257-9175
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US95-12608-31


  Query Match             51.0%;  Score 14.8;  DB 5;  Length 29;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagggcu 29
          ||:: :: :: : | || | |||||:
Db     29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT   9
US-08-632-575B-59/c
; Sequence 59, Application US/08632575B
; Patent No. 5843660
;  GENERAL INFORMATION:
;    APPLICANT:  Schumm, James W.
;    TITLE OF INVENTION:  Multiplex Amplification of
;    TITLE OF INVENTION:  Short Tandem Repeat Loci
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Promega Corporation
;      STREET:  2800 Woods Hollow Road
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53711-5399
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Mb
;      COMPUTER:  IBM compatible PC
;      OPERATING SYSTEM:  DOS, version 6.0
;      SOFTWARE:  WordPerfect 5.1 (DOS text format)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/632,575B
;      FILING DATE:  04/15/96
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/316,544
;      FILING DATE:  09/30/94
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    POSITION IN GENOME:
;      MAP POSITION:  HUMvWFA31
US-08-632-575B-59


  Query Match             51.0%;  Score 14.8;  DB 2;  Length 32;
  Best Local Similarity   42.3%;  Pred. No. 2.8e+02;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 gauucuuuuuguaagccccaagggcu 29
          ||:: :: :: : | || | |||||:
Db     29 GATTATTCTTATCATCCACTAGGGCT 4


RESULT  10
US-08-486-969-46/c
; Sequence 46, Application US/08486969
; Patent No. 5843456
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Maki, Joanne
;    TITLE OF INVENTION:  RECOMBINANT POXVIRUS - RABIES
;    TITLE OF INVENTION:  COMPOSITIONS AND COMBINATION COMPOSITIONS AND USES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
```

```
;       STREET:  530 Fifth Avenue, 25th Floor
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  United States of America
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/486,969
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Frommer, William S.
;       REGISTRATION NUMBER:  25,506
;       REFERENCE/DOCKET NUMBER:  454310-2600
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;   INFORMATION FOR SEQ ID NO:  46:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  53 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-486-969-46


  Query Match             50.3%;  Score 14.6;  DB 2;  Length 53;
  Best Local Similarity   47.6%;  Pred. No. 3.8e+02;
  Matches   10;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       9 uuuuuguaagccccaagggcu 29
           :::::|:||||  |  ||||:
Db      29 TTTTTGTAAGCTTCCCGGGCT 9


RESULT  11
US-08-410-654B-30
; Sequence 30, Application US/08410654B
; Patent No. 5833976
;  GENERAL INFORMATION:
;     APPLICANT:  Rene de Waal Malefyt
;     APPLICANT:  Di-Hwei Hsu
;     APPLICANT:  Anne O'Garra
;     APPLICANT:  Hergen Spits
;     TITLE OF INVENTION:  Use of Interleukin-10 to Treat
;     TITLE OF INVENTION:  Septic Shock
;     NUMBER OF SEQUENCES:  61
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Schering-Plough Corporation
;       STREET:  2000 Galloping Hill Road
;       CITY:  Kenilworth
;       STATE:  New Jersey
;       COUNTRY:  USA
;       ZIP:  07033
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  7.5.3
;       SOFTWARE:  Microsoft Word 5.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/410,654B
;       FILING DATE:  24-MAR-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/229,854
;       FILING DATE:  19-APR-1994
;       APPLICATION NUMBER:  US 07/926,853
;       FILING DATE:  06-AUG-1992
;       APPLICATION NUMBER:  US 07/742,129
;       FILING DATE:  06-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Foulke, Cynthia L.
;       REGISTRATION NUMBER:  32,364
;       REFERENCE/DOCKET NUMBER:  DX0221KQ1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  908-298-2987
;       TELEFAX:  908-298-5388
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  69 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (oligonucleotide)
US-08-410-654B-30


  Query Match             49.0%;  Score 14.2;  DB 2;  Length 69;
  Best Local Similarity   51.9%;  Pred. No. 6.1e+02;
  Matches   14;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 aaagauucuuuuuguaagccccaaggg 27
           ||  |: | :::  :|||| ||||| |
Db       7 AAGAATGCCTTTAATAAGCTCCAAGAG 33


RESULT  12
US-08-474-851-30
; Sequence 30, Application US/08474851
; Patent No. 5837232
;  GENERAL INFORMATION:
;     APPLICANT:  Rene de Waal Malefyt
;     APPLICANT:  Di-Hwei Hsu
;     APPLICANT:  Anne O'Garra
;     APPLICANT:  Hergen Spits
;     TITLE OF INVENTION:  Use of An Interleukin-10 Antagonist to Treat
;     TITLE OF INVENTION:  A B Cell Mediated Autoimmune Disorder
;     NUMBER OF SEQUENCES:  61
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Schering-Plough Corporation
;       STREET:  2000 Galloping Hill Road
;       CITY:  Kenilworth
;       STATE:  New Jersey
;       COUNTRY:  USA
;       ZIP:  07033
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  7.5.3
;       SOFTWARE:  Microsoft Word 6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/474,851
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/410,654
;       FILING DATE:  24-MAR-1995
;       APPLICATION NUMBER:  US 08/229,854
;       FILING DATE:  19-APR-1994
;       APPLICATION NUMBER:  US 07/926,853
;       FILING DATE:  06-AUG-1992
;       APPLICATION NUMBER:  US 07/742,129
;       FILING DATE:  06-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Foulke, Cynthia L.
;       REGISTRATION NUMBER:  32,364
;       REFERENCE/DOCKET NUMBER:  DX0221KQ1GD
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  908-298-2987
;       TELEFAX:  908-298-5388
;   INFORMATION FOR SEQ ID NO:  30:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  69 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (oligonucleotide)
US-08-474-851-30


  Query Match           49.0%;  Score 14.2;  DB 2;  Length 69;
  Best Local Similarity   51.9%;  Pred. No. 6.1e+02;
  Matches   14;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 aaagauucuuuuuguaagccccaaggg 27
          ||  |: | :::  :|||| ||||| |
Db      7 AAGAATGCCTTTAATAAGCTCCAAGAG 33


RESULT  13
US-08-481-560-30
; Sequence 30, Application US/08481560
; Patent No. 5837293
; GENERAL INFORMATION:
;    APPLICANT:  Rene de Waal Malefyt
;    APPLICANT:  Di-Hwei Hsu
;    APPLICANT:  Anne O'Garra
;    APPLICANT:  Hergen Spits
;    TITLE OF INVENTION:  Use of Interleukin-10 to Modulate
;    TITLE OF INVENTION:  Inflammation or T-Cell Mediated
;    TITLE OF INVENTION:  Immune Function
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering-Plough Corporation
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  7.5.3
;      SOFTWARE:  Microsoft Word 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,560
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  08/410,654
       FILING DATE:  24-MAR-1995
;      APPLICATION NUMBER:  US 08/229,854
;      FILING DATE:  19-APR-1994
;      APPLICATION NUMBER:  US 07/926,853
;      FILING DATE:  06-AUG-1992
;      APPLICATION NUMBER:  US 07/742,129
;      FILING DATE:  06-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Foulke, Cynthia L.
;      REGISTRATION NUMBER:  32,364
;      REFERENCE/DOCKET NUMBER:  DX0221KQ1GC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-298-2987
;      TELEFAX:  908-298-5388
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  69 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (oligonucleotide)
US-08-481-560-30


  Query Match           49.0%;  Score 14.2;  DB 2;  Length 69;
  Best Local Similarity   51.9%;  Pred. No. 6.1e+02;
  Matches   14;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 aaagauucuuuuuguaagccccaaggg 27
          ||  |: | :::  :|||| ||||| |
Db      7 AAGAATGCCTTTAATAAGCTCCAAGAG 33


RESULT  14
US-08-943-731-336/c
; Sequence 336, Application US/08943731
; Patent No. 6265157
;  GENERAL INFORMATION:
;    APPLICANT:  PROCKOP, DARWIN J.
;    APPLICANT:  SPOTILA, LORETTA D.
;    APPLICANT:  DELTAS, CONSTANTINOS D.
;    APPLICANT:  SEREDA, LARISA
;    APPLICANT:  LARSON, ANDREA W.
;    APPLICANT:  PACK, MICHAEL
;    APPLICANT:  COLIGE, ALAIN
;    APPLICANT:  EARLY, JAMES
;    APPLICANT:  KORKKO, JARMO
;    APPLICANT:  ALA-KOKKO, LEENA, et al.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;    NUMBER OF SEQUENCES:  666
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;      STREET:  FLR.
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/943,731
;      FILING DATE:  03-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/212,322
;      FILING DATE:  14-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/803,628
;      FILING DATE:  03-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOYLE LEARY Ph.D., KATHRYN
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9598-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  336:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  25 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-943-731-336


  Query Match           47.6%;  Score 13.8;  DB 4;  Length 25;
  Best Local Similarity   47.1%;  Pred. No. 7.8e+02;
  Matches    8;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 auucuuuuuguaagccc 21
```

```
          |::|: :::|: |||||
Db      22 ATTCTCTTTGTGAGCCC 6


RESULT  15
US-09-565-156A-2
; Sequence 2, Application US/09565156A
; Patent No. 6326486
; GENERAL INFORMATION:
;  APPLICANT: Hogan, James J.
;  APPLICANT:  Gordon, Patricia
;  TITLE OF INVENTION: Polynucleotide Probes for Detection and
;  TITLE OF INVENTION:  Quantitation of Bacteria in the Family Enterobacteriaceae
;  FILE REFERENCE: GP110-02.UT
;  CURRENT APPLICATION NUMBER: US/09/565,156A
;  CURRENT FILING DATE:  2000-05-03
;  PRIOR APPLICATION NUMBER: 60/132,410
;  PRIOR FILING DATE: 1999-05-03
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 41
;   TYPE: DNA
;   ORGANISM: Enterobacteriaceae
US-09-565-156A-2


  Query Match             46.9%;  Score 13.6;  DB 4;  Length 41;
  Best Local Similarity   45.0%;  Pred. No. 1e+03;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 gauucuuuuuguaagcccca 23
           | ::|: :::|:| || |||
Db       4 gcttctctttgtatgcgcca 23


Search completed: September 13, 2002, 12:37:53
Job time: 9818 sec
```

GenCore version 4.5


OM nucleic - nucleic search, using sw model

Run on: September 13, 2002, 12:36:31 ; Search time 2238.85 Seconds
(without alignments)
174.827 Million cell updates/sec

Title: US-09-310-844C-25
Perfect score: 29
Sequence: 1 aaagauucuuuuuguaagccccaagggcu 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 297742

Minimum DB seq length: 0
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| c | 1 | 18.4 | 63.4 | 70 | 9 | AA516989 | AA516989 vh89d02.r |
| c | 2 | 17.4 | 60.0 | 67 | 9 | AA708911 | AA708911 z164a10.s |
| c | 3 | 17.4 | 60.0 | 96 | 12 | AZ786158 | AZ786158 2M0031E01 |
| c | 4 | 16.8 | 57.9 | 100 | 10 | H13996 | H13996 EST00022 Ch |
| c | 5 | 16.6 | 57.2 | 70 | 9 | AI609394 | AI609394 tw93b03.x |
| c | 6 | 16.6 | 57.2 | 75 | 10 | H07686 | H07686 khgs012 BNL |
| c | 7 | 16 | 55.2 | 51 | 10 | BG361927 | BG361927 gb49d10.y |
| c | 8 | 16 | 55.2 | 73 | 10 | BG361878 | BG361878 gb46b10.y |
| c | 9 | 16 | 55.2 | 83 | 10 | BG361896 | BG361896 gb46d08.y |
| c | 10 | 16 | 55.2 | 83 | 10 | BG362080 | BG362080 gb47d02.y |
| | 11 | 16 | 55.2 | 90 | 10 | BE959933 | BE959933 601654678 |
| c | 12 | 15.8 | 54.5 | 58 | 9 | AI824019 | AI824019 wj29f03.x |
| | 13 | 15.8 | 54.5 | 76 | 12 | AZ657549 | AZ657549 1M0533L18 |
| | 14 | 15.6 | 53.8 | 37 | 9 | AI802260 | AI802260 tj36g07.x |
| c | 15 | 15.6 | 53.8 | 58 | 12 | AZ834846 | AZ834846 2M0117F18 |
| | 16 | 15.6 | 53.8 | 86 | 12 | CNS02IOD | AL199174 Tetraodon |
| c | 17 | 15.6 | 53.8 | 96 | 10 | BF730561 | BF730561 mab72e11. |
| | 18 | 15.4 | 53.1 | 49 | 10 | U44334 | U44334 ENU44334 As |
| | 19 | 15.4 | 53.1 | 77 | 12 | BH252676 | BH252676 SALK_0137 |
| c | 20 | 15.4 | 53.1 | 81 | 12 | AZ983583 | AZ983583 2M0264J16 |
| c | 21 | 15.4 | 53.1 | 86 | 10 | D25962 | D25962 HUMGS06736 |
| | 22 | 15.4 | 53.1 | 89 | 9 | AA709006 | AA709006 zf94h07.s |
| | 23 | 15.4 | 53.1 | 96 | 9 | AW641294 | AW641294 cm05g04.w |
| | 24 | 15.4 | 53.1 | 97 | 9 | AA207690 | AA207690 mv79b04.r |
| | 25 | 15.2 | 52.4 | 61 | 9 | AI318033 | AI318033 ta75g02.x |
| c | 26 | 15.2 | 52.4 | 78 | 9 | AA936218 | AA936218 on43c10.s |
| c | 27 | 15.2 | 52.4 | 98 | 9 | AW311302 | AW311302 sg35b11.y |
| c | 28 | 15 | 51.7 | 58 | 12 | B02943 | B02943 cSRL-163G2- |
| | 29 | 15 | 51.7 | 89 | 10 | BG223126 | BG223126 nah43h02. |
| c | 30 | 14.8 | 51.0 | 34 | 12 | AZ840876 | AZ840876 2M0138C08 |
| | 31 | 14.8 | 51.0 | 49 | 12 | AZ576537 | AZ576537 AST-T11C0 |
| c | 32 | 14.8 | 51.0 | 55 | 9 | AI224478 | AI224478 qx06d06.x |
| | 33 | 14.8 | 51.0 | 64 | 10 | BE636255 | BE636255 SWOvAMCAQ |
| c | 34 | 14.8 | 51.0 | 85 | 9 | AA617776 | AA617776 np99e08.s |
| c | 35 | 14.8 | 51.0 | 88 | 12 | AZ875397 | AZ875397 2M0189E08 |
| c | 36 | 14.8 | 51.0 | 95 | 9 | AI669223 | AI669223 wc13d10.x |
| | 37 | 14.6 | 50.3 | 59 | 10 | BE970792 | BE970792 601680150 |
| | 38 | 14.6 | 50.3 | 71 | 12 | AZ833202 | AZ833202 2M0115E08 |
| | 39 | 14.6 | 50.3 | 75 | 9 | AI696772 | AI696772 wc61d07.x |
| c | 40 | 14.6 | 50.3 | 82 | 12 | AZ817220 | AZ817220 2M0086M20 |
| | 41 | 14.6 | 50.3 | 85 | 10 | BF711373 | BF711373 MI-P-A1-a |
| c | 42 | 14.4 | 49.7 | 37 | 12 | AZ950243 | AZ950243 2M0214C15 |
| c | 43 | 14.4 | 49.7 | 41 | 12 | AZ598587 | AZ598587 1M0413A04 |
| c | 44 | 14.4 | 49.7 | 57 | 10 | BG362067 | BG362067 gb47b08.y |
| | 45 | 14.4 | 49.7 | 64 | 9 | AI321110 | AI321110 d4c09nm.r |

ALIGNMENTS

RESULT 1
AA516989/c
LOCUS AA516989 70 bp mRNA linear EST 14-JUL-1997
DEFINITION vh89d02.r1 Knowles Solter mouse embryonic stem cell Mus musculus cDNA clone IMAGE:894147 5' similar to TR:G187568 G187568 MG44 ;, mRNA sequence.
ACCESSION AA516989
VERSION AA516989.1 GI:2256448
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 70)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:522107
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .70
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:894147"

```
                     /clone_lib="Knowles Solter mouse embryonic stem cell"
                     /dev_stage="embryo"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; Cloned
                     unidirectionally from mRNA prepared from 800 blastocysts.
                     Primer: SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.
                     cDNAs  were cloned into the NotI/SalI sites of a pSPORT
                     vector (Life  Technologies).  "
BASE COUNT       16 a     14 c     15 g     25 t
ORIGIN


  Query Match             63.4%;  Score 18.4;  DB 9;  Length 70;
  Best Local Similarity   57.1%;  Pred. No. 7.8e+03;
  Matches   16;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 aaagauucuuuuuguaagccccaagggc 28
           | |||::|:: ::  || | ||||||||
Db      47 ACAGATTCTTCTTAGAAACACCAAGGGC 20


RESULT   2
AA708911/c
LOCUS       AA708911                 67 bp    mRNA    linear   EST 24-DEC-1997
DEFINITION  zl64a10.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:506682 3' similar to SW:RB32_HUMAN Q13637 RAS-RELATED PROTEIN
            RAB-32 ;, mRNA sequence.
ACCESSION   AA708911
VERSION     AA708911.1  GI:2718829
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .67
                     /organism="Homo sapiens"
                     /db_xref="GDB:3812701"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:506682"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT       16 a     16 c     18 g     17 t
ORIGIN


  Query Match             60.0%;  Score 17.4;  DB 9;  Length 67;
  Best Local Similarity   55.6%;  Pred. No. 1.8e+04;
  Matches   15;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3 agauucuuuuuguaagccccaagggcu 29
           |||  |:: ::|:|| |||| | |||:
Db      51 AGAGACTTCTTGTAAACCCCCAAGGCT 25


RESULT   3
AZ786158/c
LOCUS       AZ786158                 96 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0031E01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0031E01 R, DNA sequence.
ACCESSION   AZ786158
VERSION     AZ786158.1  GI:12923638
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 96)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0031  row: E  column: 01
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 96.
FEATURES             Location/Qualifiers
     source          1. .96
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0031E01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
```

```
BASE COUNT       39 a     25 c     19 g     13 t
ORIGIN


  Query Match             60.0%;  Score 17.4;  DB 12;  Length 96;
  Best Local Similarity   48.1%;  Pred. No. 1.7e+04;
  Matches   13;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      3 agauucuuuuuguaagccccaagggcu 29
          || ::|::::| | || |  |||||:
Db     54 AGTTTCTTTTTGGAGGCACTCAGGGCT 28


RESULT   4
H13996/c
LOCUS       H13996                   100 bp    mRNA    linear   EST 03-JUL-1995
DEFINITION  EST00022 Chromosome 19p12-p13.1 exon Homo sapiens cDNA clone C3-8
            5', mRNA sequence.
ACCESSION   H13996
VERSION     H13996.1  GI:888005
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   Li,Q.Y.
  TITLE     Chromosome 19p12-p13.1 exons
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Li QY
            Human Molecular Genetics
            Queen's Medical Centre
            Nottingham, NG7 2UH, UK
            Tel: 1159249924
            Fax: 1159709906
            Email: pdzqyl@pdn1.gene.nottingham.ac.uk
            Seq primer: SD2 : 5' ATC TCA GTG GTA TTT GTC AGC 3'.
FEATURES             Location/Qualifiers
     source          1. .100
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p12-p13.1"
                     /clone="C3-8"
                     /clone_lib="Chromosome 19p12-p13.1 exon"
                     /lab_host="E. coli DH5a"
                     /note="Vector: pAMP10; Exons were isolated from human
                     chromosome 19p12-p13.1 specific  cosmids from Lawrence
                     Livermore National Laboratory using a modification of  the
                     method of exon amplification (Proc. Natl. Acad. Sci. USA
                     88: 4005-4009,  1991). Amplified exons were cloned into
                     pAMP10 by uracial cloning  (GIBCOL BRL)."
BASE COUNT       22 a     35 c     24 g     19 t
ORIGIN


  Query Match             57.9%;  Score 16.8;  DB 10;  Length 100;
  Best Local Similarity   50.0%;  Pred. No. 2.8e+04;
  Matches   14;  Conservative    7;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 aagauucuuuuuguaagccccaagggcu 29
          ||||:  :::: |: ||  ||| ||||:
Db     96 AAGATGATTTTGGTTAGGTCCAGGGGCT 69


RESULT   5
AI609394/c
LOCUS       AI609394                 70 bp    mRNA    linear   EST 16-DEC-1999
DEFINITION  tw93b03.x1 NCI_CGAP_HN6 Homo sapiens cDNA clone IMAGE:2267213 3'
            similar to SW:TCPH_HUMAN Q99832 T-COMPLEX PROTEIN 1, ETA SUBUNIT ;,
            mRNA sequence.
ACCESSION   AI609394
VERSION     AI609394.1  GI:4618561
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70)
  AUTHORS   NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Dental Research,
            Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chong Heon Lee, D.D.S., Mary May, J. Silvio
            Gutkind, Ph.D., Myung Hee Park, PhD.
            cDNA Library Preparation: Stratagene, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 2028   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .70
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2267213"
                     /clone_lib="NCI_CGAP_HN6"
                     /tissue_type="normal gingiva (cell line from immortalized
                     keratinocytes)"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Cloned unidirectionally.  Primer: Oligo dT. Average
                     insert size 1.3 kb. 5' adaptor sequence: 5' AATTCGGCACGAG
                     3'                          3'     GCCGTGCTC 5' 3' adaptor
                     sequence: 5' (GA)10ACTAGTCTCGAGTTTTTTTTTTTTTTTTT 3' EcoRI
                     site appears to have been lost in a fraction of the
                     clones. Library constructed by Stratagene; available
                     through Mary May, PhD (Oral  and Pharyngeal Cancer Branch,
                     National Institute of Dental and Craniofacial Research,
                     NIH; mmay@yoda.nidr.nih.gov)."
BASE COUNT       18 a     23 c     14 g     12 t      3 others
ORIGIN


  Query Match             57.2%;  Score 16.6;  DB 9;  Length 70;
  Best Local Similarity   47.8%;  Pred. No. 3.6e+04;
  Matches   11;  Conservative    8;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagccccaagggc 28
          :: :::::|:  |||||||||  |
Db     54 TTTTTTTTGTGGGCCCCAAGGCC 32


RESULT   6
H07686/c
LOCUS       H07686                   75 bp    mRNA    linear   EST 23-JUN-1995
DEFINITION  khgs012 BNL1 Brassica napus cDNA 3', mRNA sequence.
ACCESSION   H07686
VERSION     H07686.1  GI:872508
KEYWORDS    EST.
SOURCE      rape.
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 75)
  AUTHORS   Sohn,U., Lee,C.M., Cho,K.H., Jeon,Y.H., Hahn,T.R. and Nam,H.G.
  TITLE     cDNAs from Brassica napus (rape)
  JOURNAL   Unpublished (1995)
```

```
COMMENT     Contact: Uik Sohn
            Laboratory of Molecular Biology
            Kyungpook National University
            Dept. of Genetic Eng.,Kyungpook National Univ.,Taegu 702-701,Korea
            Tel: 0539505382
            Fax: 0539555327
            Email: usohn@bh.kyungpook.ac.kr
            EST is putatively homologous to unknown gene
            Seq primer: M13 forward.
FEATURES             Location/Qualifiers
     source          1. .75
                     /organism="Brassica napus"
                     /strain="cv. Naehan"
                     /db_xref="taxon:3708"
                     /clone_lib="BNL1"
                     /lab_host="NM522"
                     /note="Vector: pT7T3D; Site_1: NotI; Site_2: EcoRI; Poly(A
                     )-mRNA was purified from the leaf of B.napus. cDNA
                     library was constructed from the mRNAs by oligo(dT)
                     priming and directionally cloned from the NotI site in
                     the vector pT7T3D (Pharmacia) to the EcoRI site. "
BASE COUNT       20 a      20 c      14 g      21 t
ORIGIN


  Query Match             57.2%;  Score 16.6;  DB 10;  Length 75;
  Best Local Similarity   56.5%;  Pred. No. 3.5e+04;
  Matches   13;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 ucuuuuuguaagccccaagggcu 29
           :|:: ::| |||| |||||| |:
Db      34 TCTTCTTGAAAGCTCCAAGGCCT 12


RESULT   7
BG361927/c
LOCUS       BG361927                 51 bp    mRNA    linear   EST 08-MAR-2001
DEFINITION  gb49d10.y1 Moss EST library PPG Physcomitrella patens cDNA clone
            PEP_SOURCE_ID: 5', mRNA sequence.
ACCESSION   BG361927
VERSION     BG361927.1  GI:13251024
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton
            ,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood
            ,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.
            , Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Physcomitrella patens"
                     /db_xref="taxon:3218"
                     /clone="PEP_SOURCE_ID:"
                     /clone_lib="Moss EST library PPG"
                     /tissue_type="gametophore: 30 day old tissue,
                     ammonium-grown"
                     /lab_host="DH10B"
                     /note="Vector: pAMP1; Construction of the cDNA library was
                     performed by Dr. W. Gregg Clark using a modification of
                     the cDNA synthesis protocol developed in the laboratory of
                     Dr. Michael Lovett by Dr. Yulia Korshunova (personal
                     communication). First polyA + RNA was isolated from total
                     gametophore RNA using oligo dT magnetic beads. Following
                     this, first strand cDNA synthesis was performed on the
                     bead-bound polyA + RNA, during which an oligonucleotide
                     anchor sequence was incorporated onto the 5'-ends of the
                     cDNA. PCR amplification was then used to synthesize the
                     second strand, to amplify the double stranded DNA, and to
                     incorporate dUTP containing sequences into the ends of the
                     double stranded cDNA. This DNA was size selected and
                     cloned into pAMP1 using the CloneAMP pAMP1 System (Life
                     Technologies, GibcoBRL) for cloning amplification products
                     by a non-restriction site dependant process. The cloning
                     was directional based on sequence asymmetry introduced at
                     the ends during PCR amplification. The 3' cDNA ends are
                     proximal to the NotI site of the multiple cloning site in
                     pAMP1. This annealing mixture was transformed into
                     chemically competent DH10B cells and selected for
                     ampicillin resistant growth. The resulting clones (about
                     330,000) were pooled to make the library."
BASE COUNT       18 a      9 c       8 g      16 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 10;  Length 51;
  Best Local Similarity   41.7%;  Pred. No. 6.3e+04;
  Matches   10;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 uucuuuuuguaagccccaagggcu 29
           :: ::::: :||| ||||||  |:
Db      27 TTTTTTTTTTAAGACCCAAGAACT 4


RESULT   8
BG361878/c
LOCUS       BG361878                 73 bp    mRNA    linear   EST 08-MAR-2001
DEFINITION  gb46b10.y1 Moss EST library PPG Physcomitrella patens cDNA clone
            PEP_SOURCE_ID: 5', mRNA sequence.
ACCESSION   BG361878
VERSION     BG361878.1  GI:13250975
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 73)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton
            ,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood
            ,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.
            , Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
```

```
    source          1. .73
                    /organism="Physcomitrella patens"
                    /db_xref="taxon:3218"
                    /clone="PEP_SOURCE_ID:"
                    /clone_lib="Moss EST library PPG"
                    /tissue_type="gametophore: 30 day old tissue,
                    ammonium-grown"
                    /lab_host="DH10B"
                    /note="Vector: pAMP1; Construction of the cDNA library was
                    performed by Dr. W. Gregg Clark using a modification of
                    the cDNA synthesis protocol developed in the laboratory of
                    Dr. Michael Lovett by Dr. Yulia Korshunova (personal
                    communication). First polyA + RNA was isolated from total
                    gametophore RNA using oligo dT magnetic beads. Following
                    this, first strand cDNA synthesis was performed on the
                    bead-bound polyA + RNA, during which an oligonucleotide
                    anchor sequence was incorporated onto the 5'-ends of the
                    cDNA. PCR amplification was then used to synthesize the
                    second strand, to amplify the double stranded DNA, and to
                    incorporate dUTP containing sequences into the ends of the
                    double stranded cDNA. This DNA was size selected and
                    cloned into pAMP1 using the CloneAMP pAMP1 System (Life
                    Technologies, GibcoBRL) for cloning amplification products
                    by a non-restriction site dependant process. The cloning
                    was  directional based on sequence asymmetry introduced at
                    the ends during PCR amplification. The 3' cDNA ends are
                    proximal to the NotI site of the  multiple cloning site in
                    pAMP1. This annealing mixture was transformed into
                    chemically competent DH10B cells and selected for
                    ampicillin resistant  growth. The resulting clones (about
                    330,000) were pooled to make the library."
BASE COUNT      24 a     15 c     13 g     21 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 10;  Length 73;
  Best Local Similarity   45.8%;  Pred. No. 5.9e+04;
  Matches   11;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagccccaagggcu 29
          :: :::::|  |||||||||  |:
Db     40 TTTTTTTTGGGAGCCCCAAGAACT 17


RESULT  9
BG361896/c
LOCUS       BG361896                 83 bp    mRNA    linear   EST 08-MAR-2001
DEFINITION  gb46d08.y1 Moss EST library PPG Physcomitrella patens cDNA clone
            PEP_SOURCE_ID: 5', mRNA sequence.
ACCESSION   BG361896
VERSION     BG361896.1  GI:13250993
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 83)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton
            ,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood
            ,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.
            , Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco.
FEATURES            Location/Qualifiers
     source         1. .83
                    /organism="Physcomitrella patens"
                    /db_xref="taxon:3218"
                    /clone="PEP_SOURCE_ID:"
                    /clone_lib="Moss EST library PPG"
                    /tissue_type="gametophore: 30 day old tissue,
                    ammonium-grown"
                    /lab_host="DH10B"
                    /note="Vector: pAMP1; Construction of the cDNA library was
                    performed by Dr. W. Gregg Clark using a modification of
                    the cDNA synthesis protocol developed in the laboratory of
                    Dr. Michael Lovett by Dr. Yulia Korshunova (personal
                    communication). First polyA + RNA was isolated from total
                    gametophore RNA using oligo dT magnetic beads. Following
                    this, first strand cDNA synthesis was performed on the
                    bead-bound polyA + RNA, during which an oligonucleotide
                    anchor sequence was incorporated onto the 5'-ends of the
                    cDNA. PCR amplification was then used to synthesize the
                    second strand, to amplify the double stranded DNA, and to
                    incorporate dUTP containing sequences into the ends of the
                    double stranded cDNA. This DNA was size selected and
                    cloned into pAMP1 using the CloneAMP pAMP1 System (Life
                    Technologies, GibcoBRL) for cloning amplification products
                    by a non-restriction site dependant process. The cloning
                    was  directional based on sequence asymmetry introduced at
                    the ends during PCR amplification. The 3' cDNA ends are
                    proximal to the NotI site of the  multiple cloning site in
                    pAMP1. This annealing mixture was transformed into
                    chemically competent DH10B cells and selected for
                    ampicillin resistant  growth. The resulting clones (about
                    330,000) were pooled to make the library."
BASE COUNT      31 a     11 c     15 g     26 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 10;  Length 83;
  Best Local Similarity   41.7%;  Pred. No. 5.7e+04;
  Matches   10;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      6 uucuuuuuguaagccccaagggcu 29
          :: :::::|:||  ||||||  |:
Db     37 TTTTTTTTGTAAATCCCAAGAACT 14


RESULT  10
BG362080/c
LOCUS       BG362080                 83 bp    mRNA    linear   EST 08-MAR-2001
DEFINITION  gb47d02.y1 Moss EST library PPG Physcomitrella patens cDNA clone
            PEP_SOURCE_ID: 5', mRNA sequence.
ACCESSION   BG362080
VERSION     BG362080.1  GI:13251177
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 83)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton
            ,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood
            ,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.
            , Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
```

```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .83
                      /organism="Physcomitrella patens"
                      /db_xref="taxon:3218"
                      /clone="PEP_SOURCE_ID:"
                      /clone_lib="Moss EST library PPG"
                      /tissue_type="gametophore: 30 day old tissue,
                      ammonium-grown"
                      /lab_host="DH10B"
                      /note="Vector: pAMP1; Construction of the cDNA library was
                      performed by Dr. W. Gregg Clark using a modification of
                      the cDNA synthesis protocol developed in the laboratory of
                      Dr. Michael Lovett by Dr. Yulia Korshunova (personal
                      communication). First polyA + RNA was isolated from total
                      gametophore RNA using oligo dT magnetic beads. Following
                      this, first strand cDNA synthesis was performed on the
                      bead-bound polyA + RNA, during which an oligonucleotide
                      anchor sequence was incorporated onto the 5'-ends of the
                      cDNA. PCR amplification was then used to synthesize the
                      second strand, to amplify the double stranded DNA, and to
                      incorporate dUTP containing sequences into the ends of the
                      double stranded cDNA. This DNA was size selected and
                      cloned into pAMP1 using the CloneAMP pAMP1 System (Life
                      Technologies, GibcoBRL) for cloning amplification products
                      by a non-restriction site dependant process. The cloning
                      was  directional based on sequence asymmetry introduced at
                      the ends during PCR amplification. The 3' cDNA ends are
                      proximal to the NotI site of the  multiple cloning site in
                      pAMP1. This annealing mixture was transformed into
                      chemically competent DH10B cells and selected for
                      ampicillin resistant  growth. The resulting clones (about
                      330,000) were pooled to make the library."
BASE COUNT       31 a     11 c     15 g     26 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 10;  Length 83;
  Best Local Similarity   41.7%;  Pred. No. 5.7e+04;
  Matches   10;  Conservative    9;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       6 uucuuuuuguaagccccaagggcu 29
           :: :::::|:||  ||||||  |:
Db      37 TTTTTTTTGTAAATCCCAAGAACT 14


RESULT  11
BE959933
LOCUS       BE959933                 90 bp    mRNA    linear   EST 14-DEC-2000
DEFINITION  601654678R1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:3839754 3',
            mRNA sequence.
ACCESSION   BE959933
VERSION     BE959933.2  GI:11776130
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 90)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     On Oct 3, 2000 this sequence version replaced gi:10570638.
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM528  row: k  column: 19
            High quality sequence stop: 72.
FEATURES              Location/Qualifiers
     source           1. .90
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3839754"
                      /clone_lib="NIH_MGC_57"
                      /tissue_type="glioblastoma"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                      SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                      Double-stranded cDNA was prepared from cell line RNA.  5'
                      and 3' adaptors were used in cloning as follows: 5'
                      adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                      sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                      (where B = A, C, or G and N = A, C,  G, or T).   Average
                      insert size 1.55 kb (range 0.9-4.0 kb).  12/15 colonies
                      contained  inserts by PCR. This library was enriched for
                      full-length clones and was constructed by Clontech
                      Laboratories (Palo Alto, CA)."
BASE COUNT       21 a     11 c     14 g     44 t
ORIGIN


  Query Match             55.2%;  Score 16;  DB 10;  Length 90;
  Best Local Similarity   50.0%;  Pred. No. 5.7e+04;
  Matches   12;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 aaagauucuuuuuguaagccccaa 24
           ||| |:: :::::| || |||| |
Db      38 AAAAATTTTTTTTGAAACCCCCCA 61


RESULT  12
AI824019/c
LOCUS       AI824019                 58 bp    mRNA    linear   EST 21-DEC-1999
DEFINITION  wj29f03.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2404253 3'
            similar to TR:O70278 O70278 MULTIPLE ENDOCRINE NEOPLASIA TYPE 1
            CANDIDATE PROTEIN NUMBER 18. ;, mRNA sequence.
ACCESSION   AI824019
VERSION     AI824019.1  GI:5444690
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 806   Std Error: 0.00
            Seq primer: -40UP from Gibco
```

```
          High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .58
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2404253"
                     /clone_lib="NCI_CGAP_Kid12"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1323912-1325831, 1471368-1472903 and
                     1492104-1493255). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT       11 a     14 c     19 g     14 t
ORIGIN


  Query Match             54.5%;  Score 15.8;  DB 9;  Length 58;
  Best Local Similarity   44.4%;  Pred. No. 7.3e+04;
  Matches   12;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 agauucuuuuuguaagccccaagggcu 29
           || :: :::::  ||| || ||| ||:
Db      56 AGCTTTTTTTTCCAAGTCCAAAGAGCT 30


RESULT  13
AZ657549
LOCUS       AZ657549                 76 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  1M0533L18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0533L18 R, DNA sequence.
ACCESSION   AZ657549
VERSION     AZ657549.1  GI:11794695
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 76)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0533  row: L  column: 18
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 76.
FEATURES             Location/Qualifiers
     source          1. .76
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0533L18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       12 a     14 c     18 g     32 t
ORIGIN


  Query Match             54.5%;  Score 15.8;  DB 12;  Length 76;
  Best Local Similarity   44.4%;  Pred. No. 6.9e+04;
  Matches   12;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 agauucuuuuuguaagccccaagggcu 29
           || ::|:::::| | || |  || ||:
Db      43 AGTTTCTTTTTGGAGGCACTCAGAGCT 69


RESULT  14
AI802260
LOCUS       AI802260                 37 bp    mRNA    linear   EST 13-DEC-1999
DEFINITION  tj36g07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2143644 3'
            similar to TR:Q41120 Q41120 HYDROXYPROLINE-RICH GLYCOPROTEIN ;,
            mRNA sequence.
ACCESSION   AI802260
VERSION     AI802260.1  GI:5367732
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 37)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 1470   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .37
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2143644"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
```

Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"
BASE COUNT 6 a 17 c 3 g 11 t
ORIGIN

Query Match 53.8%; Score 15.6; DB 9; Length 37;
Best Local Similarity 50.0%; Pred. No. 9.3e+04;
Matches 11; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aaagauucuuuuuguaagcccc 22
||| :: :::::| |||||||
Db 7 AAAATTTTTTTTTGGAAGCCCC 28

RESULT 15
AZ834846/c
LOCUS AZ834846 58 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0117F18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0117F18 R, DNA sequence.
ACCESSION AZ834846
VERSION AZ834846.1 GI:13004754
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 58)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0117 row: F column: 18
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 58.
FEATURES Location/Qualifiers
source 1. .58
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0117F18"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 15 a 13 c 16 g 14 t
ORIGIN

Query Match 53.8%; Score 15.6; DB 12; Length 58;
Best Local Similarity 54.5%; Pred. No. 8.6e+04;
Matches 12; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 4 gauucuuuuuguaagccccaag 25
| ::| :::|:|| |||||||
Db 24 GTTTCCCTTTGTAATCCCCAAG 3

Search completed: September 13, 2002, 12:36:36
Job time: 9967 sec